Artificial Neural Network and its Applications in the Energy Sector – An Overview

Damilola Elizabeth Babatunde*, Ambrose Anozie, James Omoleye

Department of Chemical Engineering, Covenant University, Otta, Ogun State, Nigeria.
*Email: damilola.babatunde@covenantuniversity.edu.ng

ABSTRACT

In order to realize the goal of optimal use of energy sources and cleaner environment at a minimal cost, researchers; field professionals; and industrialists have identified the expediency of harnessing the computational benefits provided by artificial intelligence (AI) techniques. This article provides an overview of AI, chronological blueprints of the emergence of artificial neural networks (ANNs) and some of its applications in the energy sector. This short survey reveals that despite the initial hiccups at the developmental stages of ANNs, ANN has tremendously evolved, is still evolving and have been found to be effective in handling highly complex problems even in the areas of modeling, control, and optimization, to mention a few.

Keywords: Artificial Neural Networks, Energy Sector, Optimization
JEL Classifications: Q4, P28

1. INTRODUCTION

The study of intelligence is one of the earliest disciplines (Russell and Norvig, 2016). Conjectures about the nature of intelligence have been traced back to the Greek and other Mediterranean philosophers and there has been the proposition that intelligence is related to the activity of neurons and synapses (Brunette et al., 2009). The catchphrase, "artificial intelligence (AI)" was contrived more than three decades ago, but there is no universally approved definition of the phrase. This may be due to the present abstract and immeasurable nature of intelligence (Konar, 1999). Categories of AI definition include those that relate to thought and reasoning, those that are focused on behavior, those that measure success in connection with human performance and those that evaluate against an ideal notion of intelligence termed rationality (Russell and Norvig, 2016). Psychologist and cognitive theorists are of the view that intelligence is the ability to identify the proper piece of knowledge in the right instances of decision making (Konar, 1999; Russell and Norvig, 2016). From this viewpoint, AI has been construed as the simulation of human intelligence on a machine for efficient identification and utilization of the appropriate piece of "knowledge" at a particular stage of problem-solving by the machine. Another school of thought has stated that AI is a subject that has to do with computational model with the ability to think and act rationally and this has been justified by the opinion that rationality entails all elementary characteristics of intelligence. This definition may, therefore, be taken to possess pragmatic significance. Taking rational action requires rational thinking. Successful planning is an outcome of rational reasoning because, in several instances, planning is part of a reasoning process. Possessing learning capability is integral to thinking because without the former, possession of perception is impossible. In addition, taking rational action requires obtaining adequate knowledge from real-world information (Konar, 1999).

As a specialized field, AI imitates human brain behavior and essentially relies on learning the system behavior from the obtainable/historical data of the system in order to analyze a problem domain and forecast the performance of the system being understudied. These, it does with the aid of hardware and software.

This study seeks to enunciate the relevance and current limitations of artificial neural networks (ANNs) (which is one of the foremost computational techniques) in ameliorating the challenges facing the energy sector. The rest of this paper is organized as follows: Section 2 presents the learning theories of AI, Section 3, highlights the chronological developmental stages of ANN and the basic architectures, Section 4 discusses some of the major learning mechanisms in ANNs, the implementation of ANN is presented in Section 5, Section 6 expounds the challenges facing the energy sector, Section 7 reviews the applications of ANN in resolving some of the challenges of the energy sector, and the conclusion of the study is presented in Section 8.

2. LEARNING THEORIES

Despite the successes recorded in the use of AI programs as problem solvers, learning has continued to be a challenging problem. This inadequacy appears to be severe, mainly because the ability to learn is a principal characteristic of intelligent behavior. In order to avoid unnecessary repetition of a sequence of computations for similar problems, there is a need for programs to learn from experience, analogy, examples, and instruction. While learning is a difficult research area, numerous programs have been written that indicate that it is not an impossible goal. Some approaches addressing the issue of learning in relation to AI are presented as follows.

ANNs, evolutionary computation, swarm intelligence, artificial immune systems, and fuzzy logic are some of the AI paradigms which have been used effectively to solve real-life problems (Abiyev et al., 2015; Engelbrecht, 2007; Jain and De Wilde, 2012; Zhu, 2014). Each of these paradigms originates from biological systems, has biological motivation and models biological and natural intelligence. ANNs imitate biological neural systems and permit the system to learn experiential data. Evolutionary computation is hinged on the concept of natural evolution and helps with handling imprecision (Engelbrecht, 2007; Senthilkumar, 2014). Swarm intelligence imitates the social behavior of organisms existing in swarms or colonies (Acharjya and Kauser, 2015; Engelbrecht, 2007; Senthilkumar, 2014). Artificial immune systems model the human immune system and it is primarily used for solving pattern recognition problems, executing the task of classification, and for the clustering of data (Engelbrecht, 2007). It does these in a way similar to how the natural immune system distinguishes between antigen and the cells that belong to the body through its amazing pattern matching ability. Fuzzy logic is instigated from researches on organisms' interaction with their environment and has the advantage of enabling the computer to understand natural language. Currently, there are inclinations towards developing hybrids of paradigms for the purpose of eliminating the weaknesses of specific components of a hybrid AI system by exploiting the strengths of the components (Engelbrecht, 2007).

3. ANNs APPROACH

In relation to other learning theories, the neural learning theory is the most utilized in a large number of applications. Neural learning tool which artificially simulates the behavior of a system and predicts its performance is typically regarded as an ANN. Generally, ANN is an area of AI that has the capacity of identifying the non-linear input-output relationship of a system for the purpose of diagnosing and controlling the performance of the system. It is capable of adjusting its values to correct errors from the output, and these make it an extra powerful learning tool (Ayoola et al., 2019; Makinde et al., 2012). In this session, a brief history and some concepts on the ANN are outlined.

3.1. History of ANN

Advances in biological research have made it possible to understand the natural thinking mechanism and have shown that the brain store information as a pattern. The brain is a sophisticated, nonlinear and parallel computer and has the capacity to do tasks like pattern recognition, perception, and control, swiftly than any computer. It has the ability of learning, memorization, and generalization. These capabilities of the brain form part of the motivation for studies on algorithmic modeling of biological neural systems which is termed ANNs (Engelbrecht, 2007).

In 1943, preliminary works on modeling of the functions of the brain neurons were done by physiologists, McCulloch and Pitts. The duo introduced a mathematical model of a biological neuron that is still in use today. The model has two inputs, a single output, and equal weights. The significance of the model as at then was its ability to compute any logical expression (McCulloch and Pitts, 1943). The first learning rule for the McCulloch and Pitts Neural Network which defined law for synaptic neuron learning was proposed by Donald Hebb in 1949. This law is termed the Hebbian Learning Rule. It is the most fundamental and straightforward learning rules for ANNs and deals with ways in which synapses can change their efficiencies (Hebb, 2005). Subsequently, Frank Rosenblatt, who was also a physiologist, designed the succeeding model of the neuron, called the perceptron in 1958. The perceptron was a linear system created as an effort to use neural network for character recognition and for solving problems in which the input classes were linearly separable in the input space. The ability of the perceptron model to recognize pattern was experimentally demonstrated by recognizing different simple characters (Rosenblatt, 1958). In 1960, an adaptive linear model called the ADALINE was developed by Widrow and Hoff. The model is based on the least mean square algorithm and was the first ANN to be employed in a commercial application (Widrow and Hoff, 1960). The stochastic gradient descent method for adaptive pattern classification was introduced by Amari in 1967 (Amari and Maginu, 1988).

Contrariwise, a major limitation of the perceptron which was its lack of ability to group patterns that are not linearly separable in the input space was pointed out by Minsky and Papert (Minsky and Papert, 2017). Although the mathematical proofs of the limitation of the perceptron as presented by Minsky and Papert greatly discouraged pioneering works on ANN in the 1970s, Kohonen (1972) and Anderson (1972) individually proposed the mathematical model for associative memory trained by the Hebbian Learning Rule in 1972 (Kohonen, 1972). In 1974, the backpropagation algorithm was discovered and introduced by Werbos. Backpropagation which

is a kind of gradient descent algorithm is used with ANNs for minimization and curve-fitting. This discovery is relevant to solving the limitation of the previous perception model (Werbos, 1974). In the meantime, in 1976 Grossberg studied self-organizing networks derived from the human visual systems (Grossberg, 1976). The inability of the first perceptron to deal with non-linearly separable data was not an intrinsic deficiency of the technology; it is rather an issue of scale, as the perceptron was a two-layer perceptron. In 1980, Hecht-Nielsen revealed that multilayer perceptron was able to solve nonlinear separation problems. Later on, Parker (1985) and LeCun (1985) simultaneously and independently revived backpropagation which was originally introduced by Werbos (1974). Afterward, in 1986, Hinton et al. (1986) reinvented and popularized the backpropagation algorithm (Hinton et al., 1986; LeCun, 1985; Parker, 1985; Werbos, 1974).

The works of Hinton et al. (1986) in responding to the criticism of Minsky and Papert (1969) resulted in passionate increased interest in ANN. As such, several research efforts have been successfully geared towards effective replication of the human brain behavior, the use of ANN has gone beyond simple pattern recognition problems to the very complex problems and ANN is being extensively applied on several diverse areas (Hinton et al., 1986; Minsky & Papert, 2017).

3.2. Architectures of ANN

The fundamental unit of a neural network model is the neuron. The neuron is made up of a cell body with an intrinsic nucleus; dendrites that send the external signals to the cell body; and *axons* that transfer the signals out of the cell to other cell bodies. This structure has been transformed in relation to analog computational technology as a perceptron which represents the fundamental unit of an ANN and contains processing element (comprising of summation function and transfer function); the multiple signal inputs which are linked with the perceptron through adjustable weighting elements; and the signal output(s). There is also an additional input into the perceptron known as the *bias* and is deemed to be a switching element.

The pictorial representation of the biological neuron and its analog counterpart is as shown in Figure 1.

As depicted by Figure 1, a perceptron typically accepts a number of concomitant inputs with individual input having its own proportionate weight. Each of the proportionate weight provides its respective input with the impact required of it on the summation function of the processing element. Hence, the proportionate

weights carry out a very similar task as those of the differing synaptic strengths of biological neurons. For both the synaptic weight of perceptron and synaptic strengths of biological neurons, certain inputs are made more significant than the others; this ensures that the more significant ones have more impact on the processing element when they combine to generate a neural response. The weights are adjustable coefficients within the network and they regulate the strength of the input signal as indicated by the artificial neuron. The weights indicate the connection strength of inputs and it is possible to adjust these strengths in response to different training sets, in relation to the specific architecture of a network or through its learning rules (Anderson and McNeill, 1992).

Also from equation 1, it can be seen that individual input into the perceptron is multiplied by its respective connection weight (synaptic weight). This means that the inputs are weighted. In the most basic instance, these products (weighted inputs) are just added in the summation element, and along with the bias they are passed through a transfer function to produce a result, and subsequently, the output. It is important to state that some applications do not always utilize neurons that merely sum-up, and in so doing smooth inputs. ORing, ANDing and several other functions can be built into the summation element. In other to create time-dependent networks, certain functions even integrate the input data over time (Anderson and McNeill, 1992). The transfer function (which the previously weight input and bias pass through) is also referred to as the activation function. It defines the characteristics of an artificial neuron and could be any mathematical function. The transfer function is chosen with reference to the problem that is intended to be solved by the artificial neuron in such a way that output may be enabled to satisfy more real-world interfaces. Details on the selection of activation function can be found in Kuan and White (1994) and other literature. Step function, linear function, sigmoid (non – linear) function, sine, and hyperbolic tangent are examples of the frequently used functions in the "non – linear element" of the perceptron (Anderson and McNeill, 1992; Suzuki, 2011).

Going by the pictorial description of the perceptron in Figure 1, the output from the perceptron may be mathematically defined in equation 1:

$$y_o = f\left(\sum_{i=1}^{N} w_i x_i + b_0\right) \tag{1}$$

When the bias satisfy the condition $w^T x + b_0 \geq 0$, the perceptron is activated and produces output signals; otherwise, it does not (Palit and Popovic, 2006).

Figure 1: Representation of neuron and perceptron (Palit and Popovic, 2006)
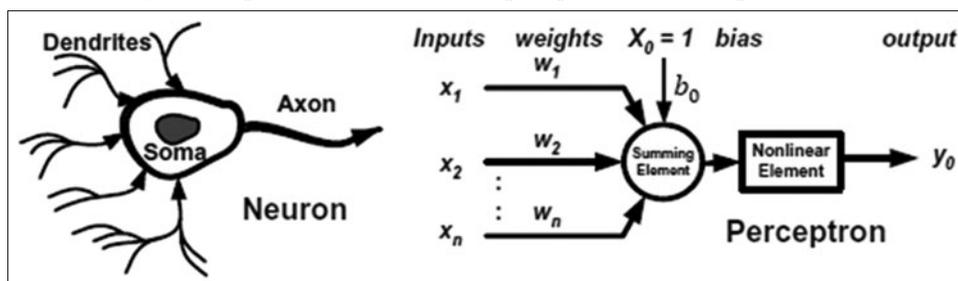
Where x_i, w_i, y_0, b_0 and w^T represents input signals, adjustable synaptic weights of the input signals, output signal, bias, transpose of the synaptic weights, respectively.

While the operational principles and basic set of rules of artificial neuron seem so ordinary, the maximum capacities and computing power of artificial neuron models can be harnessed when they are interconnected. When two or several artificial neurons are combined, an ANN is derived; and ANNs utilize the fact that complexity can emerge from just a few fundamental and simple rules (Suzuki, 2011). As a result of processing information in their elementary units in a non-linear, distributed, parallel and local way, ANNs possess the capability to solve complex practical problems that lone artificial neurons cannot solve (Suzuki, 2011).

Architecture, topology, and graph of an ANN have been used interchangeably to mean the way a number of neurons are organized or positioned with respect to one another; the arrangements which are principally structured by directing the synaptic connections of the neurons. In an effort to maximally harness the pros of the mathematical complexity obtainable from the interlinking of artificial neurons and in order not to make the system unnecessarily complex and unmanageable, the individual artificial neurons are not just connected randomly (Suzuki, 2011). A sizeable number of previous researches have been directed towards standardization of ANN topography and several standardized" topographies of ANNs have emerged from those researches. These predefined topographies make problem-solving easier, faster and more efficient. Distinctive topographies are suitable for addressing different categories of problems. The topology of the ANN has to be selected and fine-tuned once the problem type to be solved has been identified. Fine-tuning the ANN topology itself and its parameters is a precondition for getting the network ready for use (Suzuki, 2011). The interlinking of artificial neurons can be carried out in numerous ways and this results in several possible topologies. The possible topologies are grouped into two basic classes which are the feedforward and the recurrent topologies. In the feedforward topologies, information flows only unidirectionally from the inputs to the outputs. On the contrary, some of the information not only flow from inputs to outputs but also flow in the reverse direction in recurrent topologies. In addition to the basic classes of topologies, neurons making up a network are grouped into layers for ease of handling and mathematical description of the ANNs (Suzuki, 2011). These layers are: the input layer (which receives data, signals, features, or measurements from the external environment); the hidden/invisible/intermediate layer (comprising neurons which extract patterns connected with the process being analyzed and execute most of the internal processing of the network); and thirdly, the output layer (comprising of neurons which generates and presents the ultimate outputs of the network, which are outcomes of the processing implemented by the neurons in the hidden layers) (da Silva et al., 2017; Suzuki, 2011). Figure 2 is a simplified illustration of an ANN.

3.3. Models of ANN

3.3.1. Feedforward ANN

ANNs with feedforward topology are known as feedforward ANNs. The only condition for feedforward neural network is the unidirectional flow of information from inputs to outputs without any back-loops. As stated in the previous session, input layer neurons receive information from the environment. This information is transferred weighted (by the connection weights between the input and hidden neurons) to the neurons in the intermediate layer. The intermediate layer neurons then transform the corresponding input signals by the activation function in each of the neurons, to create neural signals. The neural signals generated by the intermediate layer are also weighted by the connection weights between the hidden neurons and the output neurons and are forwarded to the neurons in the output layer. These weighted signals are then transformed by the activation function in each corresponding output neuron. Conversely, the signals that determine the actions required are thus, produced by the output neurons. All information emanating from the neurons in a "lower" layer is processed concurrently (and not successively) by the neurons in an "upper" layer and there is a constant term which is referred to as the bias term in each of the activation functions which adds flexibility to the responses of the intermediate and output neurons.

The most basic feedforward ANN is a single perceptron, which has the ability to learn linear separable problems only. However,

Figure 2: Illustration of simple artificial neural network (Suzuki, 2011)
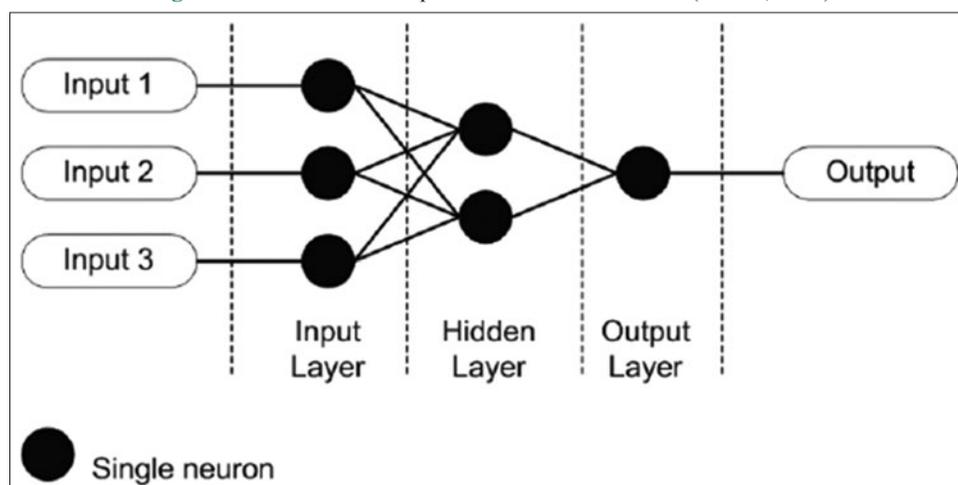
there are no restrictions on the number of layers, the number of neurons in each layer, the type of transfer function employed in each neuron or the number of links between the neurons in feedforward ANN. It is worthy to note that, even simple multilayer feedforward ANNs as in the case of the feedforward ANN with one hidden layer (with the input, hidden and output layers having three, two and one neurons respectively) shown in Figure 3 and described using equations (2) through (9), result in relatively long mathematical description in which the network's parameter optimization cannot be done manually. Computers and specialized software are used for mathematically description, building and parameter optimization of all types of ANNs for practical use (Suzuki, 2011). Another notable characteristic of feedforward ANN is the inability to memorize information because of the lack of back – loops. The feedforward neural networks treat all sample data as new even if the sample data have temporary dependence on previous signals (Kuan and White, 1994).

$$n_1 = F_1\left(w_1 x_1 + b_1\right) \tag{2}$$

$$n_2 = F_2\left(w_2 x_2 + b_2\right) \tag{3}$$

$$n_3 = F_2\left(w_2 x_2 + b_2\right) \tag{4}$$

$$n_4 = F_3\left(w_3 x_3 + b_3\right) \tag{5}$$

$$m_1 = F_4\left(q_1 n_1 + q_2 n_2 + b_4\right) \tag{6}$$

$$m_2 = F_5\left(q_3 n_3 + q_4 n_4 + b_5\right) \tag{7}$$

$$y = F_6\left(r_1 m_1 + r_2 m_2 + b_6\right) \tag{8}$$

$$y = F_6 \left[\begin{array}{l} r_1\left(F_4\left[q_1 F_1\left[w_1 x_1 + b_1\right] + q_2 F_2\left[w_2 x_2 + b_2\right]\right] + b_4\right) + \ldots + \\ r_2\left(F_5\left[q_3 F_2\left[w_2 x_2 + b_2\right] + q_4 F_3\left[w_3 x_3 + b_3\right]\right] + b_5\right) + b_6 \end{array} \right] \tag{9}$$

Where x, n, m denotes input signals into the input, hidden and output layers, respectively. y represents the output signal from the network. w, q, r symbolizes the adjustable weights of signal inputs to the input, hidden and output layers, respectively. F and b represent the transfer functions and biases, respectively.

3.3.2. Radial basis function (RBF) neural network

RBF networks are a distinctive category of feedforward neural networks that have a three-layer architecture. The network is simply connected to its environment by the input layer. The invisible layer comprises a number of neurons, and these neurons nonlinearly transform the input parameters with a RBF like the Gaussian function, the thin plate spline function, etc. The output layer is linear and functions as a summation unit. Figure 4 shows the representative structure of an RBF neural network.

Some of the pros of RBF networks are the ease of design, effective generalization, high tolerance to input noises and online learning ability. From the point of generalization, RBF networks are able to respond adequately to patterns not previously used for training (Yu et al., 2011).

3.3.3. Kohonen neural network

Kohonen neural networks are a variant of feedforward neural networks that utilize an unsupervised learning approach. They organize the output neurons into a spatial map by means of a process that can be termed as self-organization; as such they are otherwise known as Kohonen self-organizing maps (SOM) (Akerkar and Sajja, 2010). A Kohonen network is made of two layers of processing units which are the input and output layers. The network has no hidden layer. As input pattern is being fed into the network, there is competition between the neurons in the output layer, and the neuron which has it's incoming connection weights nearest (with respect to the Euclidean distance) to the input pattern normally wins. Therefore, as the input is being introduced to the network, individual output neuron calculates its nearness or matches score to the input pattern. The output that is nearest to the input pattern "wins" and earns the entitlement to have its connection weights modified. The connection weights are shifted in the direction of the input pattern by a factor estimated by a learning rate parameter. The topological mapping created by the Kohonen network is achieved not only by adjusting the winner's

Figure 3: Simple multi-layer feedforward artificial neural network (Suzuki, 2011)
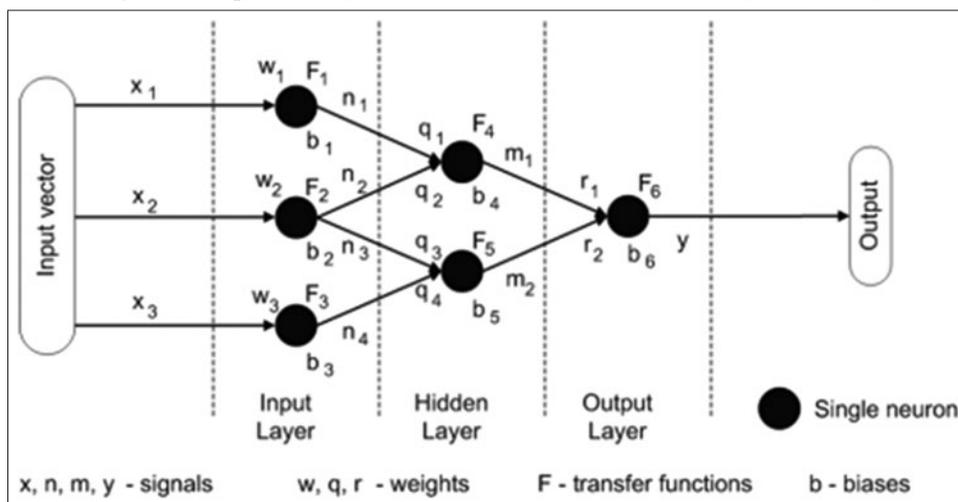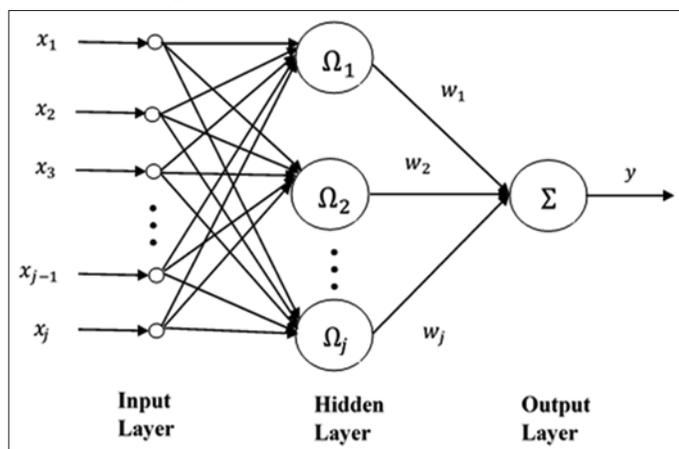
Figure 4: Typical structure of an radial basis function network

weights but also by modifying the weights of the adjacent output neurons in the vicinity to the winner. This ensures that not only are the weights of winners are modified, the entire neighborhood of output neurons is also drawn nearer to the input pattern. Beginning with the randomized weight values, the output neurons gradually align themselves in a way that when an input pattern is introduced, a neighborhood of neurons responds. As training progresses, the size of the neighborhood radiating out from the winning unit is decreased. The rate of learning will decline as training continues, and in certain executions, the rate of learning declines with the distance from the winning output neuron.

3.3.4. Recurrent neural networks

As stated in the preceding subsection, the feedforward neural network is unable to memorize data. When the sample data have temporary dependence on previous signals as in the case of learning to predict future elements of a time series, the feedforward ANN has to be expanded with a memory mechanism to take care of the dynamic patterns. This, in addition to feedforward connections, will require feedback loops to preserve past information in the form of the information processing state. The ANN which possesses recurrent topology is known as a recurrent neural network.

The recurrent neural network has no limitations with respect to back loops and information is transmitted both in the forward and backward directions. The flow of information in both directions generates an internal state of the network which permits internal feedback and exhibition of dynamic temporal behaviour. The simplest topology of the recurrent ANN is referred to as a fully recurrent artificial network; it is a network in which each of the fundamental elements (artificial neurons) is directly linked to all other basic elements in every direction. Hopfield, Elman, Jordan, bi-directional are some unique examples of recurrent ANNs.

4. THE ANN LEARNING MECHANISM

Training an ANNs is among the most vital matters associated with the operational use of their potentials. Once the type and structure of the network and the transfer function have been selected for a specific application, the next thing is to train the network so it can learn to satisfactorily respond to inputs. The training of a selected

architecture of ANN involves adjustment of the synaptic weights and thresholds of its neurons by application of a set of ordered steps. The process of adjustment is termed learning algorithm and its purpose is to tune the network in order to obtain outputs that are close to the targeted values. It is essential that an appropriate learning strategy is chosen for training the network (Szymczyk and Szymczyk, 2015). Some of the distinguishing factors between learning mechanisms are the training data types accessible to the learner, the order and method by which training data is received and the test data used to appraise the learning algorithm. The commonly used learning mechanisms which include supervised learning, unsupervised learning and reinforced learning are elucidated in this section.

4.1. Supervised Learning

Supervised learning is a technique in which the learner accepts the expected outputs for a particular set of input signals and uses it as training data to predict all hidden points (Mohri et al., 2018). For this learning strategy, individual training samples comprises the input signals and their respective outputs. Therefore, an input-output data table which typifies the process and its behavior is required. This input-output data table is known as an attribute/value table. The attribute table provides the information which the neural structure uses in developing a hypothesis about the system it is learning (da Silva et al., 2017). In this way, the network behaves as if there is a coach teaching it the appropriate response to every input. To obtain the desired response to every input made available to the network, the learning algorithm compares the outputs produced to the desired outputs. Based on the disparity observed, the network continues to modify the synaptic weights and thresholds using the adjustment procedure. In view of the purposes of generalizing solutions, the network is deemed "trained" when the discrepancy falls within a satisfactory range of values. Hence, supervised learning is a representative example of pure inductive inference, in which the network's free variables are attuned based on foreknowledge of the anticipated outputs for the system being investigated. This learning strategy is most commonly often used for classification, regression, and ranking problems (Mohri et al., 2018).

4.2. Unsupervised Learning

Unsupervised learning is a technique in which the learner is provided with entirely unlabeled training data for making predictions of all hidden points (Mohri et al., 2018). In other words, the network receives inputs but are not provided with desired outputs. With this learning strategy, the parameters of the ANNs are set with reference to the data given and a cost function which is to be minimized. The cost function can be any function and it is determined by the task formulation. In unsupervised learning, the focus is to determine how the data is organized (Suzuki, 2011). The features to be used in grouping the input data is determined by the system itself. This is usually referred to as self-organization or adaption. The network has to organize itself when particularities exist among the elements that make up the whole sample set, and in addition, subsets (or clusters) possessing similarities also have to be identified (da Silva et al., 2017; Suzuki, 2011). For the identified clusters within the network to be taken into account, the learning algorithm modifies the synaptic weights and thresholds of the network. On the other hand,

the network designer can stipulate beforehand, the highest possible number of these probable clusters, based on the understanding of the problem (da Silva et al., 2017). Quantitative evaluation of a learner's performance can be difficult because generally, no labeled example is available for the learner (Mohri et al., 2018). Unsupervised learning finds practical applications mostly within the domain of estimation problems such as clustering, dimensionality reduction, compression, statistical modeling, filtering and blind source separation (Mohri et al., 2018; Suzuki, 2011).

Although unsupervised learning is currently not completely understood as a full-blown learning strategy, the strategy can be used to carry out certain initial characterization on inputs and its ability to adapt to the environment makes it an excellent promising learning strategy that can be found practically suitable in real-life situations which rarely present exact training sets and for situations where the unexpected aspect to life has to be prepared for. An example of such situations is military action where new warfare techniques and novel weaponries might be encountered (Anderson and McNeill, 1992).

4.3. Reinforcement Learning

Reinforcement learning is the study of planning and learning in a setup where a learner actively interacts with the environment to realize a particular goal. It is a technique that is aimed at programming the learner by reward and punishment without requiring specification on how a task is to be achieved (Kaelbling et al., 1996). Reinforcement learning focuses on how to map situations to actions in such a way that the numerical reward signal is maximized. Unlike most machine learning strategies, reinforcement learning is not provided with what actions to take but by itself has to find out which actions give the most reward by trying them. This is however associated with difficult computational obstacles. In some cases, actions may not only affect the immediate reward but also affect subsequent situations, and consequentially, all succeeding rewards (Sutton and Barto, 2018). Without stating which action is to be taken, the reinforcement signal which the environment provides is simply given to an intelligent agent for it to take good and appropriate action. The intelligent agent, therefore, relies on former experience actions and outcomes. To get an enormous reward, a reinforcement learner has to prefer previous actions that it has tested and proven to be effectual in yielding reward. In addition to that, it must test actions that it has not chosen previously. This implies that it is necessary for the agent to leverage what it previously knows so as to obtain a reward. In addition, it must also explore for better action choices to be made in the future (Qiang and Zhongli, 2011).

Reinforcement learning methods are taken to be a variant of supervised learning methods, as they constantly evaluate the departure of the network's response from the desired output (Sutton and Barto, 2018). However, in reinforcement learning, the learning process is heuristic because the response for a particular input can either be satisfactory or unsatisfactory and if the response is satisfactory, the behavioral condition associated with the system is reinforced (rewarded) by a gradual increment of the synaptic weights and thresholds. The heuristic search and deferred reward are the principal differentiating characteristics of reinforcement learning (Sutton and Barto, 2018).

A number of learning algorithms employed by reinforcement learning are centered on stochastic approaches that choose the modification actions probabilistically, by taking into account a finite set of probable solutions that can be rewarded if they have likelihoods of producing satisfying outcomes. In the course of the training process, the probabilities associated with action adjustment are modified to improve network performance. The reinforcement learning strategy has been found useful in solving control optimization problems. Usually, control optimization problems have to do with identifying the best action in every state visited by the system for the purpose of optimizing certain objective functions. When a system has a very huge amount of states (>1000) and has a complex stochastic structure, which is not amenable to closed-form analysis, the reinforcement learning technique is usually used (da Silva et al., 2017).

5. IMPLEMENTATION OF ANN

Several software packages and program codes are currently available for the implementation of the underlying principles of diverse mythologies of ANNs to forecast the performance of any type of system. Generally, developing a valid and effective ANN model requires an appropriate choice of activation functions. More often than not, once the choice of activation is made, the unknown parameters (connection weights) will be estimated and subsequently, the right number of invisible units will be determined. The network must properly be trained to ensure it learns the connection weights (Kuan and White, 1994).

The number of hidden units determines the complexity of the network; hence the complexity has to be regularized by finding the appropriate number of hidden units. When the training sample is fixed, a large number of hidden units may result in data overfitting. To prevent data overfit while seeking to enhance approximation capability during the implementation of ANN models, network complexity may be regularized by applying a standard selection criterion. Bayesian information criterion (BIC) and predictive stochastic complexity (PSC) are some of the selection criteria that are well known. The BIC has two terms; one of the terms has to do with model fitness while the second term penalizes model complexity. With these two terms, the BIC is appropriate for the regularization of network complexity. The PSC approach gives an average of squared prediction errors. For a given number of inputs, the network having the lowest BIC or PSC provides the desired number of invisible units (Kuan and White, 1994).

A sufficient understanding of activation functions is required for making the appropriate choice. Activation functions are mathematical functions that convert the input to an ANN to output. There are several activation functions that can be used in ANN. The most basic is the linear activation function. It gives a direct relationship between the input and output and it is denoted by the expression in equation 10:

$$y = f(x) \tag{10}$$

The linear activation function is normally used for the output layer activation function. Most of the problems being addressed using

ANN are nonlinear in nature. To achieve nonlinearity, high degree polynomial functions which are also known as nonlinear functions are required because they provide the complexity factor needed to make ANN a true universal function approximator. Since 2015, the rectified linear unit (ReLU) has been the most used activation function (equation 11).

$$. \ f(x) = 0 \ when \ x < 0$$
$$. \qquad x \ when \geq 0 \qquad\qquad (11)$$

ReLU is a simple condition and has merits over the other functions. Ciaburro and Venkateswaran (2017) offered extensive information on some other functions such as unit step activation function, sigmoid and hyperbolic tangent. They posit that the chosen activation function should generally be differential and not cause gradients to vanish; it should be simple and fast in processing; and should not be zero centered because of the nonlinear and complex problems that ANNs are meant to handle (Ciaburro and Venkateswaran, 2017).

6. CHALLENGES FACING THE ENERGY SECTOR

The main objectives of the power generating industry are to supply electricity at minimal cost and to constantly render quality service (Babatunde et al., 2019). These objectives are rarely fully accomplished. Some of the factors limiting the achievement of these objectives are the inevitable constant change in load, rapidly increasing energy demand, changes in weather conditions, and the need to achieve a cleaner environment. (Bourguet and Antsaklis, 1994). The conventional approach of generating electrical energy from non-electrical ones typically involve combustion of fossil-based fuels such as natural gas and coals and the process of generating electricity from these fuels basically requires boiler, turbine, and generator; and by extension, some auxiliary equipment which work together with the main units to meet the megawatt demand (Bourguet and Antsaklis, 1994). Like every other industrial process where production target is rarely static, power generating plants often have to operate at off-design or part-load conditions because of the fluctuating power demands or short supply of fuels. Changes in megawatt demand (load variation) have been reported to result in changes in the efficiency of the power plant (Karakurt, 2017; Nakamura and Toyota, 1988). This is a result of the changes in process parameters that inherently accompany changes in megawatt demand. Of these process parameters, change in superheated steam temperature has the largest effect on the efficiency of steam power plants and effective control has to be in place to maintain the superheated steam at the optimum temperature (Nakamura and Toyota, 1988). Other key parameters also have to be precisely controlled and maintained at desired values required for the target performance of power plants. In time past when the operation of power plants was over a baseload, single input - single output (SISO) feedforward control strategies were adequate for the plants. However, as demand grows, and fuel crises ensue, and new environmental restrictions emerge, the control strategies which were hitherto adequate have to be replaced with more efficient control strategies. In addition to these, the task of control is complex because of the high coupling

among the process variables and the nonlinear nature of the process (Bourguet and Antsaklis, 1994). Asides from effective control mechanisms, effective load forecast models to predict the needs of the electrical grid is a necessity.

According to Sorrell (2015), most analysts have pointed out that improving efficiency and reducing energy demand largely contribute to obtaining a cleaner environment. This assertion, however, seems to contradict the globally acclaimed direct link between wealth and energy demand (Sorrell, 2015). Sorrell (2015) also posited that the relationship between improved efficiency and energy demand reduction is far from been forthright because of the divergent interpretation of improved energy efficiency and as such, reduction in energy demand can be difficult than is generally presumed. In further establishing the author's position on the multifacetedness of energy demand reduction as a subject matter, the relationships between energy demand and economic growth, energy efficiency, energy market, energy policy, and sociotechnical systems were expounded. In the process of integrating various perspectives on energy demand reduction, the weight ascribed to each factor that affects the demand for energy differs. From an economic perspective, Sorrel (2015) pointed out that demand for energy will not reduce if energy prices do not rise and if there are no policies to reduce the economic hindrances to improved energy efficiency. In addition to the issue of favorable policy, interventions that foster energy-efficient choices and support for new, energy-efficient technologies at various stages of innovation chain are also required for energy demand reduction. Taking these further, optimization techniques offer a systematic approach to establish necessary trade-offs when designing systems that will run efficiently and generate lesser pollutants at a minimal cost. When inefficiencies are observed in any section of a plant, optimization serves as a tool for maximizing or minimizing process parameters as a corrective measure to increase efficiency. Multi-variable optimizations, which consider important interactions between subsystems are veritable tools for achieving optimum performance of plants and they rely on reliable models of the plants. Notably, the interaction between subsystems making up a plant is complex, and models depicting these interactions are highly nonlinear in nature. The development of such comprehensive models from the first principle is highly cumbersome as they require deep knowledge of several physical sciences and engineering principles, which may not be handy.

Energy from renewable sources is known to be environment-friendly and inexpensive (Babatunde et al., 2018; Babatunde et al., 2019; Babatunde et al., 2019). Although renewable sources of energy may not be able to generate sufficient energy to fully replace fossil fuels in the short term, their complementary role in supplying/meeting the energy demand is vital to the sustenance of modern civilization (Babatunde et al., 2019). The numerous challenges associated with the incorporation of renewable energy plants to the conventional power grid, however, limit their contribution to the dependable and viable energy supply, which is critical and inevitable to the sustenance of contemporary civilization (Babatunde et al., 2019). The dependence on weather conditions and the intermittency of these renewable energy sources is one of the foremost challenges. These attributes make matching supply and load cumbersome, thereby upsetting the stability of the

electrical network. It is therefore pertinent that accurate model that can forecast the intermittency of these renewable energy sources be developed in order to have an economical access to the energy sources, effectively manage the entire electrical network and reduce the negative cost impact of intermittency on the network (Ferrero et al., 2019; Fouilloy et al., 2018; Notton et al., 2018). Although there are properly developed fundamental physical models for the individual components of all types of renewable energy generation systems, the direct characterization of these systems using closed mathematical terms is not practical. This is because of the complexities which arise from the combination of these components. In addition to the intermittent nature of renewable energy sources, some challenges encountered in integrating renewable energy plants to conventional grids are linked with the reliability of the systems generating the energy. The issue of reliability is also a leading cause of power outages in standalone conventional power plants. An effective fault diagnosis system to detect failures in electrical components of both standalone and integrated electrical systems is pertinent to achieving improved reliability (Bourguet and Antsaklis, 1994; Ferrero et al., 2019).

Inferences drawn from literature summarized indicate the multifaceted and complex nature of the problems facing the energy sector, and obtaining a viable solution to these problems requires tools that are able to adequately handle the complex relationship between the variables associated with the problems. In a literature survey conducted by Ferrero et al. (2019), it was pointed out that ANN is recommended for use when it is desired to generate new knowledge that is otherwise difficult to obtain; improve forecasting accuracy with a wide range of variable; when documentation of activities and data from variables, and replication of results with a high quality depends on good procedures and information systems and when results which are flexible and dynamically adaptable in model implementations are preferred over exact and very accurate results. It is also very useful in deciphering complex relationships between variables, and it possesses a huge tolerance to data uncertainty. However, ANN is not without its inherent limitations. A major limitation to the use of ANN techniques is the reliance on a vast amount of quality historical data which may not be easily accessible or obtainable at a fair cost. (Ferrero et al., 2019; Navarro and Bennun, 2014). In illustrative examples by Navarro and Bennun, (2014), the implied limitation that ANN cannot produce results that are as accurate as those obtained by statistical methods when stochastic events are involved, seems to have been contradicted by the research outcomes of the prediction of flow discharge of a river in the work of Fereydooni et al. (2012). Furthermore, some research works in certain areas of applications of ANN, indicate that the integration of statistical theories and methods with ANN techniques have improved the prediction accuracy when stochastic data are analyzed, thereby helping to overcome this limitation in those areas of applications (Ling et al., 2016; Mosavi et al., 2018; Saglietti et al., 2018; Samli, 2012; Tran et al., 2019; Wang et al., 2012).

7. APPLICATION OF ANN IN THE ENERGY SECTOR

The power industry has been experiencing tremendous developments due to the use of AI techniques such as genetic algorithm, fuzzy logic, ANN, and expert systems. These techniques all present cost-effective developments to the industry. Within the framework of AI – machine learning techniques, ANN models have produced good outcomes for real-time estimations, particularly when dynamic changes in conditions of the environment is to be learned as a major factor to increase the accuracy of forecasting (Ferrero et al., 2019; Fouilloy et al., 2018). In light of the advantages of ANN over other techniques, some of its applications in the areas of need in the energy sector are presented in this section.

7.1. Identification, Modeling, and Prediction

In an overview of ANNs in the electric power industry presented by Bourguet and Antsaklis (1994), projects considered were classified into applications of ANN for power plants and applications for power systems. Some of the applications of ANN in power plants highlighted by the Bourguet and Antsaklis include identification and modeling; sensor validation; monitoring and fault diagnosis; and control. In power systems, they considered the use of ANN for static and dynamic security assessment; transient stability assessment; identification, modeling and prediction; control; and fault diagnosis. In all of the cases they reviewed, feedforward ANN and backpropagation training algorithm were predominantly used. Having outlined successful applications of ANN in a representative number of research projects in the industry, the duo established the strength and applicability of ANN technology to the industry and submitted that power systems computing with neural nets are taken to be one of the fastest-growing field in power system engineering. Acir (2013) studied the use of ANN in performance evaluation of a coal-fired thermal power plant. The ANN model for forecasting the exergetic efficiency of the plant was developed using PHYTICA software with a backpropagation neural network based on the Levenberg–Marquardt algorithm. Environmental temperature, condenser pressure, and steam pressure were taken as input into the network to serve as the performance factors. Out of the 27 values of the actual dataset, 2-third was used for the network training while a third was randomly selected to test the performance of the trained network. Having explored six different network structures (3-3-1, 3-4-1, 3-4-5-1, 3-5-4-1, 3-3-4-5-1 and 3-4-3-4-1), Acir discovered that the 3-3-4-5-1 was the best for the plant under study and the novel method will aid both simplified and quick calculation of exergetic performance by researchers. Chokshi et al. (2018), explored the use of ANN for estimating the thermal performance (both energetic and exegetic) of a 210 MW coal-fired thermal plant as it operates round – the – clock at various loads. In the study which was geared toward predicting the instantaneous performance and the effect of critical parameters on the KWU designed thermal plant, Chokshi et al. (2018) setup sixty-five ANN models by varying the number of neurons or the spread constants in the hidden layer of each of the five ANN architectures they considered. The models were trained using actual datasets of the plant and were subsequently used to forecast the performance of a different KWU plant with equal capacity. Having evaluated the error in performance prediction using five different approaches, the generalized regression network with spread constant of one, was selected and recommended as an appropriate model to forecast the performance of a 210 MW KWU power plant under varied operational load. Furthermore, it was established that the

recommended model is useful for the determination of optimal values of various critical parameters for the optimal performance of the plant at varied operational load.

Literature on the use of ANN for modeling the highly nonlinear relationships among variables connected with emissions resulting from combustion of fuels that generate the heat needed to raise the enthalpy of water in boilers are also available (Ilamathi et al., 2013; Ronquillo-Lomeli et al., 2018; Saleh et al., 2015; Sun et al., 2013; Yusoff and Aziz, 2009). Using ANN, Yusoff and Aziz (2009) modeled emissions from the boiler of a palm oil mill that runs on shells and fibres which are waste from oil palm processing. Fifteen parameters which include shell and fibre flowrates, ambient temperature, combustion temperature, flue gas temperature, steam temperature, and pressure were taken as input variables for the model. These parameters were measured from different locations in the plant. The output variables of the model were emissions of CO, NO_x, SO_2 and PM. In comparison to the predicted outputs from the Multiple Linear Regression technique, the ANN model exhibited more flexibility and accuracy in emission prediction. From the viewpoint of Yusoff and Aziz (2009), the ANN models developed can effectively predict and monitor pollutant emissions from several processes, serve the same purpose as continuous emission monitoring system (CEMs) and in terms of operation and maintenance, less expensive than CEMs.

7.2. Energy Demand/Load Forecasting

Load forecast is an important reference point for effective power planning required for proper management of energy, smooth operation and stability of power systems (Babatunde et al., 2018). It helps to adequately plan operations, optimize power generation, minimize the cost associated with the dispatch of load and serves as a guide for planning infrastructural investment required for power system expansion (Samuel et al., 2017). A survey of the uses of ANN in power industry, especially in the areas of load forecasting, economic load dispatch, and security assessment was conducted by Mohatram et al. (2011). In comparison with other techniques, the survey reveals that ANN is very quick regardless of the intricacies of the problems; possesses an extraordinary ability for on-line processing and classification; possess massively parallel distributed structure and ability to learn; and has the capability for implicit nonlinear modeling and automatic filtering of system data. On the flip side, determination of the number of hidden layers, number of neurons in the hidden layer and the time-consuming nature of the process of finding the optimal configuration of ANNs are some of the challenging issues of ANN mentioned by the researchers. Using ANN, Adepoju et al. (2007) performed a short-term load forecast which gives the load demand experienced by a utility company in Nigeria based on "an hour ahead of time." Past load data from the Company for a period of 1 month were acquired and pre-processed before using for the training of the network and the load forecast. The network consisted of the input layer, an intermediate layer, and the output layer. There were five input variables to the network and this corresponds to five neurons for the input layer. A single intermediate layer was selected for the network's topology in order to avoid the complexities and tardiness associated with multiple intermediate layers. In addition to this, the number of neurons

in the intermediate layer was varied from five to eleven, before eventually employing eleven neurons in that layer (because it gave a better model behavior). The variation of the number of neurons was to prevent the loss of generalizing ability caused by excessive neurons and the inability to learn the characteristics of a dataset caused by insufficient neurons in the intermediate layer. The network had only one output parameter, which corresponds to a single neuron in the output layer. Sigmoid transfer function and linear transfer function was used in the intermediate and output layer respectively. Supervised training of the network was done using data of the first 2 weeks of the month and by supplying the input values with the corresponding expected output value. With an absolute mean error of 2.54% recorded in load forecast for the next 1 h when the network was tested with another week's load data, the ability of the ANN to effectively forecast short-term load was confirmed. The results indicate that the neural network has been able to learn the nonlinear correlation between the past load data presented to it during the training period and effective forecasts have been made based on this.

Samuel et al. (2017) carried out a comparative study on the medium-term load forecast abilities of ANN and regression models using load data from the power station of a teaching-research institution in Nigeria. The entire dataset spanning about 184 days was partitioned into training, validation and forecast sub-datasets. The neural network toolbox in MATLAB was used for the design and configuration of the ANN model. A multi-layer perceptron (MLP) network with an input layer, two intermediate layers, and an output layer was considered in the study. For the ANN architecture considered, the input layer consisted of four neurons and used the "tansig" function, while the output layer had only a neuron and applied the 'purelin' function. Using the inputs and the expected output, the training of the network was done with a variant of the backpropagation learning algorithm known as the gradient descent algorithm. Validation of the model was carried out by comparing its outputs to those of the validation datasets and the parameters of the network were tuned until the best load forecasting ANN model was obtained. Taking the root mean square error (RMSE) and the mean absolute percentage error (MAPE) as the basis of comparison, the load forecast made by the ANN model was deemed more accurate than those of the regression models (cubic, compound-growth and linear). Whereas the regression methods are able to give instant forecast outcomes because they simply require direct mathematical calculations, the ANN model gives faster load forecast once the training has been completed (Samuel et al., 2017).

A survey of methods used in conducting long-term electric load forecasts was carried out by Panda et al. (2017). In the survey, theoretical features of three parametric methods (trend analysis, end-use technique, and econometric technique) and four AI centered techniques (ANN, fuzzy logic model, wavelet networks and genetic algorithm) were highlighted. In addition to discussing the fundamental principles of these methods, the survey reveals that the ANN method was the most used computational technique for long-term load forecasts in the scenarios considered. ANN-based models yielded long-term electric load forecasts which are close to those of the real data. Furthermore, ANN models were able

to overcome the inadequacies of parametric techniques in handling nonlinear problems. With lower forecast error relatively to those of the parametric methods and other computational intelligence methods studied, and the capability to produce satisfactory outcomes even when available historical data is low, the ANN model was taken to be a superior technique for long-term load forecast. Many other works on the use of ANN for electricity demand forecasts are available in the literature (Bozkurt et al., 2017; Buhari and Adamu, 2012; Hernández et al., 2014).

7.3. Monitoring and Fault Diagnosis

Monitoring and diagnosis of faults in rotating equipment such as pumps and centrifuge are vital tasks for the smooth running of most industrial plants, particularly in power generating plants because of the role they play in the prevention of line break-down, enhanced efficiency, reliability, and safety of the plant. Continuous monitoring has been recommended for early detection of faults in order to prevent negative consequences and this can be laborious, time-intensive and error-prone when handled by human operators (Azadeh et al., 2013). To that end, there has been increased use of AI and other machine learning techniques in the development of monitoring and fault diagnosis schemes for major plant equipment that are prone to fault. Of these machine learning techniques that are being explored by researchers, ANNs are the most prevalent. Azadeh et al. (2013) proposed a unique flexible algorithm for the classification of centrifugal pump conditions. In the proposition of the algorithm, two types of faults were classified based on six standard pointers (flow, temperature, suction pressure, discharge pressure, velocity, and vibration) and the capability of ANN model, pure support vector classification (SVC) and hybrid SVC were compared in both normal and noisy environments. Based on the percentage of fault types predicted correctly, the outcomes of the study indicate that support vector machine (SVM) centered approaches performed better than the ANN method. However, the ANN approach gave better results in comparison with some highly appraised methods of classification, such as K-nearest neighbors (KNN) and decision tree. A nonparametric statistical test which is known as McNemar's test that was conducted at a 5% significance level to determine which model significantly outperformed the other further indicates that no statistically significant difference exists between the SVC models and ANN model. In addition, Azadeh et al. (2013) discovered that although the ANN approach was not robust in a noisy environment at all times, it still outperformed KNN and decision tree and the calculated performance of the proposed flexible algorithm fall within the approved limit. Fast and Palme (2010) applied ANN to develop an online condition monitoring and diagnosis system for a combined heat and power plant (CHP). The online monitoring and diagnosis system for the plant involves the creation of ANN models for the component equipment making up the plant. These models were integrated on the server of the power generation information manager in the computer system of the CHP, the graphical user interface was accessed through the workstations linked with the server and training of the models were done with operating data from the plant. The outcomes of the study indicate ease of integration with the computer system of the plant and accurate prediction by ANN model. As a contribution towards minimization of environmental impact and increased reliability, availability and maintainability of

existing power plants, Fast et al. (2009) developed ANN model for an industrial gas turbine and demonstrated the multi-utility of the model. The multi-utility ANN model which was built using multi-layer feed-forward network was trained with operational data by means of back-propagation. The outcome of the study revealed the possibility of predicting the operative and performance parameters (such as identification of anti-icing mode) of gas turbine with good precision at various local environmental temperature. According to Fast et al. (2009), the multi-utility ANN model developed is suitable for offline performance simulation of the gas turbine; online monitoring of the condition of the turbine for prompt fault detection and prevention of degradation and; can also serve the purpose of sensor validation. With the aid of a graphical user interface, the model was able to give instantaneous estimations of the gas turbine performance. Asides from being able to give good results of extrapolated data beyond the range of the training data, the chief importance of the ANN model was the effectiveness in detecting compressor fouling which helped optimized the wash interval of the compressor (Fast et al., 2009). Having posited that a monitoring system consists of several elements, and divers factors determine the ultimate diagnostic precision, Loboda (2016) conducted investigations based on the gas path of gas turbine to underpin the author's position on precision in monitoring of systems. Faults of components were simulated using gas turbine thermodynamic model; diagnostic information was extracted from raw operating data by computing the deviance of each variable being monitored from the reference values; and a number of ANN-based networks alongside KNN and support vector network (SVN) were considered for fault classification, fault diagnosis, and turbine monitoring. The outcome of the study includes proposition and proving of methods for increased accuracy of diagnosis; and recommendations on selecting and adapting the networks for various diagnostic tasks. More literature on monitoring and fault diagnosis in gas turbines, and other energy systems and components are vastly available (Dybkowski and Klimkowski, 2019; Palmé et al., 2011; Patel and Shah, 2018; Simani and Fantuzzi, 2000).

Timely discovery and diagnosis of boiler trips have been prescribed for continued safe operation and as an antidote to unnecessary hike in the operating cost of thermal power plants. This assertion formed the basis of the research on an intelligent cautionary scheme for boiler tube leak trips conducted by Singh et al. (2017). This research which focused on obtaining an intelligent scheme capable of optimal prediction of boiler trips considered three actual cases of thermal power plant boiler tube leak trip. Two intelligent cautionary schemes were proposed; the first was a pure ANN-based scheme which was trained using extreme learning machine (ELM) algorithm and the other scheme was a hybridized system of ANN and genetic algorithm which was also trained using ELM method. The proposed schemes were validated with actual faulty data and in all the three cases considered, the prompt discovery of boiler tube leak trips was achieved. The pure ANN-based intelligence scheme detected leak trip earlier than the hybrid intelligent scheme. On the other hand, the hybrid intelligent scheme was taken to be the more reliable scheme because of the optimization capability of the genetic algorithm. Furthermore, the Root Mean Square Error in all the trip cases were lesser and was therefore considered to outperform the

pure ANN cautionary scheme. Alnaimi and Al-Kayiem (2011) also worked on fault diagnosis in a boiler of a power plant by focusing on monitoring the superheater through an AI scheme. The research involved the development of ANN models for detection of faults and diagnosis. With regards to superheater monitoring, the most impelling parameters contributing to low-temperature trip in boiler were determined by studying the boiler behavior and thirty-two parameters were selected. One invisible-layer and two invisible-layer ANN architectures were explored and feed-forward back-propagation was basically used for the neural network training. Since the deployed training algorithm has several variations based on the minimization algorithm used for minimizing error in the error estimator, four kinds of variants were studied. The outcomes of the exploration show that superheater low temperature can be promptly and accurately detected with satisfactory performance using the models developed.

7.4. Control

In implementing control, an accurate model of the process is a requirement. Data-actuated models are handy in representing a particular process when it is difficult to develop knowledge-actuated simulation models if an ample amount of data depicting the process is at disposal (Solomatine et al., 2009). Major data-driven modeling paradigm includes neural networks, fuzzy-based and genetic algorithm systems. ANN has been recognized as an invaluable tool that can support the effective development of data-actuated models for even extremely nonlinear and multivariable systems (Al Seyab and Cao, 2008; Cao et al., 2008). ANN can learn complex functional relations for a system from the input and output data of the system (Liptak, 2018).

In a comparative study done on tuning a PID controller by Srinivas et al. (2014), intelligent methods have been reported to give a prompt response with lesser peak overshoot and integral square error when juxtaposed with Ziegler-Nichols and Cohen-Coon methods (Srinivas et al., 2014). Out of the intelligent methods explored by these researchers, ANN-tuning turned out to be the best. Al Fayiz (2017), designed, tested and assessed a model based on an adaptive neural network algorithm for the purpose of controlling a nuclear power plant. The results of the assessment indicate that in contrast to traditional methods of control, the understudied adaptive algorithm is capable of substantially improving the control quality of the automatic power control system. Manke and Tembhurne (2012), also developed and tested an ANN-based drum level controller for boilers of thermal power plants. The aim of developing the ANN-based controller was to considerably minimize the cases of failure of boiler tubes caused by tube overheating (when water level suddenly drops), difficulty in moisture-steam separation & reduced boiler efficiency (caused by rise in water level), and the associated boiler & turbine trips; plant shutdown, forfeiture of power generation and loss of income. The controller prototype developed has a simple structure and is based on a simple algorithm. The learning samples were empirical data from a 500 MW coal-fired power plant and was trained using backpropagation technique. When applied to control the drum level in a thermal power plant, the developed ANN controller was able to overcome the limitations of the existing PI controller. The better performance and prospect of handling complex process

control-related problems make the ANN-based controller highly promising for process-control applications. Anead et al. (2018), Also conducted research on the use of ANN-based control for boiler damage prevention. ANN controller was designed with operational data obtained from an industrial steam power plant. The data consisting of parameters that have been identified to have a significant influence on the performance of the boiler were trained using backpropagation algorithm. The purpose of the training was to enable the controller to make right decision for the control of the boiler. The off-line control of thermal variables was implemented using simulink in MATLAB and the outcome indicate that the highest deviation of the predicted input variables from the experimental optimum value was below 0.01.

7.5. Optimization

Optimization techniques centered on ANNs, fuzzy logic, genetic algorithms, and some other computational intelligence approaches have been identified as being capable of producing good results in adverse situations where data is not precise, noisy and inconsistent (Kumar and Karimi, 2015). In a bid to improve on clean and efficient use of coal for electricity generation, Ilamathi et al. (2012) combined ANN with simulating annealing (SA) to obtain optimum combustion parameters that minimize the emission of oxides of nitrogen (NOx) from coal combustion. In the study, ANN served as an effective tool for mapping the nonlinear input-output relationship, which yielded the fitness function between the input operating variables and NOx emission. This fitness function provided by the ANN served as the objective function which SA utilized in searching for the optimum operating parameters to meet the required low level of NOx emission. The implementation of both the ANN modeling and simulated annealing was carried out with MATLAB R2011a. Data for the modeling and optimization study was obtained from the experiment in which a number of operating parameters were varied and their effect on the level of NOx emissions of a 210 MW thermal power plant operated at full load was investigated. The ANN which deciphered the nonlinear relationship between the operating variables and NOx emitted had 9 neurons (corresponding to 9 parameters) at the input layer, 10 neurons and 10 biases at the intermediate layer, and 1 neuron and 1 bias at the output layer. The "transig" function and "purelin" function were used to compute the sum of the weighted inputs and bias(es) at the intermediate and output layer respectively. To derive the fitness function, the network was trained using feedforward backpropagation. A comparison of the NOx emissions level predicted by the trained network to those of the measured ones shows that the network performed considerably well. The optimum solution which SA obtained by using the fitness function derived from ANN reveals that a combination of ANN and SA can be used to generate realistic operational conditions that optimize coal combustion with respect to NOx emission reduction.

As concerns are increasingly being raised about the devastating effects of greenhouse gases emitted into the environment from various sources and anthropogenic activities, considerable research efforts are being geared towards reducing the contribution of landfills to the non-CO_2 greenhouse gases released into the environment. This is mostly achieved by harnessing the biomethane obtainable from landfill gas as a form of renewable energy for

power generation (Denwigwe et al., 2019; Ighravwe and Babatunde, 2019). A research on the optimization of biomethane production from a biogas plant located near a landfill, using the combination of ANN model and genetic algorithm model was carried out by Qdais et al. (2010). The plant's operative data over a period of 177 days were acquired and preprocessed in preparation for the study. Using the preprocessed data, a multi-layer ANN having two hidden layers, each with 25 neurons and sigmoid activation function was used to determine the nonlinear impacts of reactor temperature, total solids, total volatile solids and pH on the quantity of biomethane produced by the plant's digester. Backpropagation training of the network was implemented in MATLAB. The validation of the ANN model designed was done using a set of 50 days' operative data that were not previously used for the training of the network. The network was deemed to have been adequately trained, as the ANN model effectively predicted the production of biomethane with a correlation coefficient of 0.87. The output of the ANN model was used in computing the values of the fitness function for the genetic algorithm routine. Hence, the optimum combination of parameters (based on the four input variables) for operating the digester of the biogas plant was identified through the integration of the ANN model and the GA model. These optimum conditions led to a 6.9% increase in the yield of methane as compared with the maximum yield obtained with the current operative parameters. Kana et al. (2012) also studied the use of ANN and genetic algorithm for the optimized production of biogas from a combination of five different substrates. Data from twenty-five mini-pilot biogas fermentations were acquired and preprocessed for the study. Using the Neurosolution software, the ANN was structured in a way that it had 5 input nodes, a single hidden layer with 2 neurons, and 1 output node; and the sigmoid transfer function was adopted in the processing elements. Eighty percent of the biogas fermentation data was used for the training of the neural network using the backpropagation approach. Having been successfully trained with the concentration datasets of the five substrates, validation of the ANN model for prediction of the digester's performance index was done with the remaining data. After the validation of the model, it was used as the objective function for the process of genetic algorithm optimization. The implementation of the optimization using genetic algorithm was done with Pro-optimizer software. This yielded an optimum profile which resulted in an 8.64% increase in biogas production and the initialization of biogas production on day 3 of fermentation as against day 8 in the non-optimized system. In other words, the experimental evaluation of the results obtained from the modeling and optimization study confirmed that the integration of ANN and GA models can efficiently be used to optimize the non-linear biogas production process.

8. CONCLUSION

The energy sector is faced with multifarious challenges that can no longer be adequately tackled with conventional approaches and tools only. The computational strengths of several artificial intelligent schemes have been proven to be veritable tools in providing new and effective ways of handling a lot of the existing and emerging challenges faced by the sector. By dint of the adaptive nature of the ANN learning procedures, and in relation to other artificial computational schemes, ANN is deemed an extra-powerful tool and the most explored technique in solving problems in the energy sector.

The ANN has gone through several developmental stages and its development is still ongoing. The highlights provided in this survey unveils its robustness in effectively handing problems extending from modeling, prediction, load forecast, fault detection, monitoring to control, in different areas of the energy sector. In cases where its lone ability is insufficient, the combinative strength drawn from its integration with other computational intelligence schemes cannot be downplayed as in the scenarios of optimization cited in this review.

9. ACKNOWLEDGMENT

REFERENCES

Abiyev, R.H., Aliev, R., Kaynak, O., Turksen, I.B., Bonfig, K.W. (2015), Fusion of Computational Intelligence Techniques and Their Practical Applications. Computational Intelligence and Neuroscience. Call for Papers.

Acharjya, D.P., Kauser, A.P. (2015), Swarm intelligence in solving bio-inspired computing problems: Reviews, perspectives, and challenges. In: Handbook of Research on Swarm Intelligence in Engineering. Hong Kong: IGI Global. p74-98.

Acir, A. (2013), Application of artificial neural network to exergy performance analysis of coal fired thermal power plant. International Journal of Exergy, 12(3), 362-379.

Adepoju, G.A., Ogunjuyigbe, S.O.A., Alawode, K.O. (2007), Application of neural network to load forecasting in Nigerian electrical power system. The Pacific Journal of Science and Technology, 8(1), 68-72.

Akerkar, R., Sajja, P. (2010), Knowledge-based Systems. Sudbury: Jones and Bartlett Publishers.

Al Fayiz, M.H. (2017), Adaptive neural network algorithm for power control in nuclear power plants. Journal of Physics Conference Series, 781(1), 12-52.

Al Seyab, R.K., Cao, Y. (2008), Differential recurrent neural network based predictive control. Computers and Chemical Engineering, 32(7), 1533-1545.

Alnaimi, F.B.I., Al-Kayiem, H.H. (2011), Artificial intelligent system for steam boiler diagnosis based on superheater monitoring. Journal of Applied Sciences, 11(9), 1566-1572.

Amari, S.I., Maginu, K. (1988), Statistical neurodynamics of associative memory. Neural Networks, 1(1), 63-73.

Anderson, J.A. (1972), A simple neural network generating an interactive memory. Mathematical Biosciences, 14(3-4), 197-220.

Anderson, D., McNeill, G. (1992), Artificial neural networks technology. Kaman Sciences Corporation, 258(6), 1-83.

Anead, H.S., Sultan, K.F., Abd-Kadhum, R.J. (2018), Evaluation and improvement performance of a boiler in a thermal power plant using artificial neural network. Engineering and Technology Journal, 36, 656-663.

Ayoola, A.A., Hymore, F.K., Omonhinmin, C.A., Olawole, O.C., Fayomi, O.S.I., Babatunde, D., Fagbiele, O. (2019), Analysis of waste groundnut oil biodiesel production using response surface methodology and artificial neural network. Chemical Data Collections, 22, 100238.

Azadeh, A., Saberi, M., Kazem, A., Ebrahimipour, V., Nourmohammadzadeh,

A., Saberi, Z. (2013), A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyper-parameters optimization. Applied Soft Computing, 13(3), 1478-1485.

Babatunde, D.E., Babatunde, O.M., Akinbulire, T.O., Oluseyi, P.O. (2018), Hybrid energy systems model with the inclusion of energy efficiency measures: A rural application perspective. International Journal of Energy Economics and Policy, 8(4), 310-317.

Babatunde, O.M., Adedoja, O.S., Babatunde, D.E., Henry, D.I. (2019), Off-grid hybrid renewable energy system for rural healthcare centers: A case study in Nigeria. Energy Science and Engineering, 7, 1-18.

Babatunde, O.M., Denwigwe, I.H., Adedoja, O.S., Babatunde, D.E., Gbadamosi, S.L. (2019), Harnessing renewable energy for sustainable agricultural applications. International Journal of Energy Economics and Policy, 9(5), 308-315.

Babatunde, O.M., Munda, J.L., Hamam, Y. (2018), Generation Expansion Planning: A Survey. Africa: IEEE PES/IAS Power. p307-312.

Babatunde, O.M., Munda, J.L., Hamam, Y. (2019), A comprehensive state-of-the-art survey on power generation expansion planning with intermittent renewable energy source and energy storage. International Journal of Energy Research, 43, 6078-6107.

Bourguet, R.E., Antsaklis, P.J. (1994), Artificial neural networks in electric power industry. ISIS, 94, 7.

Bozkurt, Ö.Ö., Biricik, G., Taycsi, Z.C. (2017), Artificial neural network and SARIMA based models for power load forecasting in Turkish electricity market. PLoS One, 12(4), e0175915.

Brunette, E.S., Flemmer, R.C., Flemmer, C.L. (2009), A Review of Artificial Intelligence. Wellington, New Zealand: 4th International Conference on Autonomous Robots and Agents. p385-392.

Buhari, M., Adamu, S.S. (2012), Short-term Load Forecasting Using Artificial Neural Network. Hong Kong: International Multi Conference of Engineers and Computer Scientists. p14-16.

Cao, Y., Jin, Y., Kowalczykiewicz, M., Sendhoff, B. (2008), Prediction of Convergence Dynamics of Design Performance Using Differential Recurrent Neural Networks. 2008 IEEE International Joint Conference on Neural Networks. Hong Kong: IEEE World Congress on Computational Intelligence). p528-533.

Chokshi, R.B., Chavda, N.K., Patel, A.D. (2018), Prediction of performance of coal-based KWU designed thermal power plants using an artificial neural network. International Journal of Applied Engineering Research, 13(5), 3093-3110.

Ciaburro, G., Venkateswaran, B. (2017), Neural Networks with R: Smart Models Using CNN, RNN, Deep Learning, and Artificial Intelligence Principles. Birmingham: Packt Publishing Ltd.

da Silva, I.N., Spatti, D.H., Flauzino, R.A., Liboni, L.H.B., dos Reis Alves, S.F. (2017), Artificial neural network architectures and training processes. In: Artificial Neural Networks. New York: Springer. p21-28.

Denwigwe, I.H., Babatunde, O.M., Babatunde, D.E., Akintunde, T.J., Akinbulire, T.O. (2019), A technical review on methods and tools for evaluation of energy footprints, impact on buildings and environment. In: Energy Footprints of the Bio-refinery, Hotel, and Building Sectors. Singapore: Springer. p47-81.

Dybkowski, M., Klimkowski, K. (2019), Artificial neural network application for current sensors fault detection in the vector controlled induction motor drive. Sensors, 19(3), 571.

Engelbrecht, A.P. (2007), Computational Intelligence: An Introduction. Hoboken: John Wiley & Sons.

Fast, M., Assadi, M., De, S. (2009), Development and multi-utility of an ANN model for an industrial gas turbine. Applied Energy, 86(1), 9-17.

Fast, M., Palme, T. (2010), Application of artificial neural networks to the condition monitoring and diagnosis of a combined heat and power plant. Energy, 35(2), 1114-1120.

Fereydooni, M., Rahnemaei, M., Babazadeh, H., Sedghi, H., Elhami, M.R. (2012), Comparison of artificial neural networks and stochastic models in river discharge forecasting,(Case study: Ghara-Aghaj River, Fars province, Iran). African Journal of Agricultural Research, 7(40), 5446-5458.

Ferrero, B.J., Fernández, J.F.G., Polo, F.O., Márquez, A.C. (2019), A review of the use of artificial neural network models for energy and reliability prediction. A study of the solar PV, hydraulic and wind energy sources. Applied Sciences, 9(9), 1844.

Fouilloy, A., Voyant, C., Notton, G., Motte, F., Paoli, C., Nivet, M.L., Duchaud, J.L. (2018), Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability. Energy, 165, 620-629.

Grossberg, S. (1976), Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. Biological Cybernetics, 23(3), 121-134.

Hebb, D.O. (2005), The Organization of Behavior: A Neuropsychological Theory. New Delhi: Psychology Press.

Hernández, L., Baladrón, C., Aguiar, J., Calavia, L., Carro, B., Sánchez-Esguevillas, A., Lloret, J. (2014), Artificial neural network for short-term load forecasting in distribution systems. Energies, 7(3), 1576-1598.

Hinton, G.E., McClelland, J.L., Rumelhart, D.E., Rnmelhart, D.E. (1986), Parallel Distributed Processing: Explorations in The Microstructure of Cognition. Cambridge: MIT Press.

Ighravwe, D.E., Babatunde, D.E. (2019), Evaluation of landfill gas plant siting problem: A multi-criteria approach. Environmental Health Engineering and Management Journal, 6(1), 1-10.

Ilamathi, P., Selladurai, V., Balamurugan, K. (2012), Predictive modelling and optimization of nitrogen oxides emission in coal power plant using artificial neural network and simulated annealing. IAES International Journal of Artificial Intelligence, 1(1), 11.

Ilamathi, P., Selladurai, V., Balamurugan, K. (2013), A novel approach for modelling of NOx emission reduction in a tangentially fired coal boiler. International Journal of Oil, Gas and Coal Technology, 6(4), 449-461.

Jain, L., De Wilde, P. (2012), Practical Applications of Computational Intelligence Techniques. Vol. 16. London: Springer Science and Business Media.

Kaelbling, L.P., Littman, M.L., Moore, A.W. (1996), Reinforcement learning: A survey. Journal of Artificial Intelligence Research, 4, 237-285.

Kana, E.B.G., Oloke, J.K., Lateef, A., Adesiyan, M.O. (2012), Modeling and optimization of biogas production on saw dust and other co-substrates using artificial neural network and genetic algorithm. Renewable Energy, 46, 276-281.

Karakurt, A.S. (2017), Performance analysis of a steam turbine power plant at part load conditions. Journal of Thermal Engineering, 3(2), 1121-1128.

Kohonen, T. (1972), Correlation matrix memories. IEEE Transactions on Computers, 100(4), 353-359.

Konar, A. (1999), Artificial Intelligence and Soft Computing: Behavioral and Cognitive Modeling of The Human Brain. Boca Raton: CRC Press.

Kuan, C.M., White, H. (1994), Artificial neural networks: An econometric perspective. Econometric Reviews, 13(1), 1-91.

Kumar, U., Karimi, M.N. (2015), Comparison and suitability assessment of available optimization techniques for optimizing waste heat recovery in industries. YMCAUST International Journal of Research, 3(2), 1-9.

LeCun, Y. (1985), A learning scheme for asymmetric threshold networks. Proceedings of Cognitiva, 85(537), 599-604.

Ling, H., Samarasinghe, S., Kulasiri, D. (2016), Stochastic neural networks for modelling random processes from observed data. In: Artificial Neural Network Modelling. Berlin: Springer. p83-107.

Liptak, B.G. (2018), Instrument Engineers' Handbook, Volume Two: Process Control and Optimization. London: CRC Press.

Loboda, I. (2016), Neural networks for gas turbine diagnosis. In: Artificial Neural Networks? Models and Applications. Cham: InTech.

Makinde, F.A., Ako, C.T., Orodu, O.D., Asuquo, I.U. (2012), Prediction of crude oil viscosity using feed-forward back-propagation neural network (FFBPNN). Petroleum and Coal, 54(2), 120-131.

Manke, P., Tembhurne, S. (2012), Application of back propagation neural network to drum level control in thermal power plants. International Journal of Computer Science Issues, 9(2), 520.

McCulloch, W.S., Pitts, W. (1943), A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics, 5(4), 115-133.

Minsky, M., Papert, S.A. (2017), Perceptrons: An Introduction to Computational Geometry. Cambridge: MIT Press.

Mohatram, M., Tewari, P., Shahjahan. (2011), Applications of artificial neural networks in electric power industry: A review. International Journal of Electrical Engineering, 4(2), 161-171.

Mohri, M., Rostamizadeh, A., Talwalkar, A. (2018), Foundations of Machine Learning. Cambridge: MIT Press.

Mosavi, A., Ozturk, P., Chau, K. (2018), Flood prediction using machine learning models: Literature review. Water, 10(11), 1536.

Nakamura, H., Toyota, Y. (1988), Statistical identification and optimal control of thermal power plants. Annals of the Institute of Statistical Mathematics, 40(1), 1-28.

Navarro, H., Leonardo, B. (2014), Descriptive examples of the limitations of artificial neural networks applied to the analysis of independent stochastic data. International Journal of Computer Engineering and Technology, 5(5), 40-42.

Notton, G., Nivet, M.L., Voyant, C., Paoli, C., Darras, C., Motte, F., Fouilloy, A. (2018), Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting. Renewable and Sustainable Energy Reviews, 87, 96-105.

Palit, A.K., Popovic, D. (2006), Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications. Berlin: Springer Science and Business Media.

Palmé, T., Fast, M., Thern, M. (2011), Gas turbine sensor validation through classification with artificial neural networks. Applied Energy, 88(11), 3898-3904.

Panda, S.K., Mohanty, N.S., Jagadev, A.K. (2017), Long term electrical load forecasting: An empirical study across techniques and domains. Indian Journal of Science and Technology, 10(26), 1-16.

Parker, D. (1985), Learning-logic: Casting the cortex of the human brain in silicon. Cambridge: Center for Computational Reserch.

Patel, H.R., Shah, V.A. (2018), Fault detection and diagnosis methods in power generation plants-the Indian power generation sector perspective: An introductory review. PDPU Journal of Energy and Management, 2(2), 31-49.

Qdais, H.A., Hani, K.B., Shatnawi, N. (2010), Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm. Resources, Conservation and Recycling, 54(6), 359-363.

Qiang, W., Zhongli, Z. (2011), Reinforcement Learning Model, Algorithms and Its Application. Jilin, China: International Conference on Mechatronic Science, Electric Engineering and Computer (MEC). p1143-1146.

Ronquillo-Lomeli, G., Herrera-Ruiz, G., Rios-Moreno, J., Ramirez-Maya, I., Trejo-Perea, M. (2018), Total suspended particle emissions modelling in an industrial boiler. Energies, 11(11), 3097.

Rosenblatt, F. (1958), The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65(6), 386.

Russell, S.J., Norvig, P. (2016), Artificial Intelligence: A Modern Approach. Malaysia: Pearson Education Limited.

Saglietti, L., Gerace, F., Ingrosso, A., Baldassi, C., Zecchina, R. (2018), From statistical inference to a differential learning rule for stochastic neural networks. Interface Focus, 8(6), 20180033.

Saleh, C., Leuveano, C., Achmad, R., Ab Rahman, M.N., Md Deros, B., Dzakiyullah, N.R. (2015), Prediction of CO_2 emissions using an artificial neural network: The case of the sugar industry. Advanced Science Letters, 21(10), 3079-3083.

Samli, R. (2012), Stochastic neural networks and their solutions to optimisation problems. International Journal of Electronics, Mechanical and Mechatronics Engineering, 2(3), 293-297.

Samuel, I.A., Adetiba, E., Odigwe, I.A., Felly-Njoku, F.C. (2017), A comparative study of regression analysis and artificial neural network methods for medium-term load forecasting. Indian Journal of Science and Technology, 10(10), 7.

Senthilkumar, S. (2014), Practical applications of swarm intelligence and evolutionary computation. International Journal of Swarm Intelligence and Evolutionary Computation, 3(1), 1000e106.

Simani, S., Fantuzzi, C. (2000), Fault diagnosis in power plant using neural networks. Information Sciences, 127(3-4), 125-136.

Singh, D., Ismail, F.B., Nasif, M.S. (2017), Hybrid Intelligent Warning System for Boiler tube Leak Trips. MATEC Web of Conferences. p131.

Solomatine, D., See, L.M., Abrahart, R.J. (2009), Data-driven modelling: Concepts, approaches and experiences. In: Practical Hydroinformatics. Berlin: Springer. p17-30.

Sorrell, S. (2015), Reducing energy demand: A review of issues, challenges and approaches. Renewable and Sustainable Energy Reviews, 47, 74-82.

Srinivas, P., Lakshmi, K.V., Kumar, V.N. (2014), A comparison of PID controller tuning methods for three tank level process. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 3(1), 6810-6820.

Sun, B.M., Wang, D.H., Yang, B., Zhang, S.H., Kong, L.Y. (2013), Prediction Model for the boiler NOx emission with material properties based on the artificial neural network. Advanced Materials Research, 676, 40-45.

Sutton, R.S., Barto, A.G. (2018), Reinforcement Learning: An Introduction. Cambridge: MIT Press.

Suzuki, K. (2011), Artificial Neural Networks-methodological Advances and Biomedical Applications. London: InTechOpen.

Szymczyk, P., Szymczyk, M. (2015), Supervised learning laplace transform artificial neural networks and using it for automatic classification of geological structure. Neurocomputing, 154, 70-76.

Tran, A.D., Duc, D.T., Pham, V.S. (2019), Improved rainfall prediction using combined pre-processing methods and feed-forward neural networks. Journal, 2(1), 65-83.

Wang, J., Pan, H., Liu, F. (2012), Forecasting crude oil price and stock price by jump stochastic time effective neural network model. Journal of Applied Mathematics, 5-6, 1-15.

Werbos, P. (1974), Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Ph. D. Dissertation. Cambridge: Harvard University.

Widrow, B., Hoff, M.E. (1960), Adaptive Switching Circuits. New York: IRE WESCON Convention Record.

Yu, H., Xie, T., Paszczyñski, S., Wilamowski, B.M. (2011), Advantages of radial basis function networks for dynamic system design. IEEE Transactions on Industrial Electronics, 58(12), 5438-5450.

Yusoff, A.R., Aziz, I.A. (2009), Predicting boiler emission by using artificial neural networks. Jurnal Teknologi, 50(1), 15-28.

Zhu, X. (2014), Computational intelligence techniques and applications. In: Computational Intelligence Techniques in Earth and Environmental Sciences. Berlin: Springer. p3-26.