2.2.2.1 Association Rule Mining

Association rule mining (ARM) refers to algorithms aiming at discovering rules (if-then statements) that describe large portions of data [30], for example, people that likes **X** tend to also like **Y**. Another general example would be algorithms whose goal is to find common values of the variables in the data that appear with a large degree of frequency. The first algorithms proposed for association rules were the AIS algorithm in 1993 [31] and the set-oriented mining algorithm [32] just one year later. These two strategies still have relevance nowadays as a basis for further developments in the area. Another popular line of research for ARM was the one following the Apriori and AprioriTid algorithms, which were designed to avoid the effort wastage of counting candidate item sets that are known to be infrequent [33]. Taking these approaches as a base, several algorithms have been proposed to overcome limitations, such as scanning the whole data set many times or handling data sets with a large number of item sets. To avoid providing a huge list of algorithms, we can summarize the techniques by grouping them in, for example, frequent pattern tree algorithms, rapid association rule mining algorithms, and frequent pattern mining algorithms, to name a few [34, 35]. The ARM-based methods can also be classified into weighted ARM, quantitative ARM, and temporal ARM, in general a sense.

2.2.2.2 Clustering-Based Methods

Clustering methods (sometimes also referred to as data segmentation methods) have different goals related to grouping or segmenting the objects or items into groups (also called "clusters") in which the objects belonging to a given cluster are more closely related given a certain metric than objects assigned to a different group [26]. One of the issues that clustering analysis presents is that groups are formed based on a notion of similarity, defined by a given criterion that directly impacts the way groups are formed. For instance, one of the most popular algorithms for clustering, the k-means algorithm [36, 37], groups the data in function of the Euclidean distance to the cluster center. Another drawback of this algorithm is related to the number of clusters that are formed, being a parameter selected by the user. This drawback was later addressed with automatic clustering algorithms that find a fair number of clusters without a specific definition of the user. Examples of these algorithms are the mean-shift clustering algorithms [38], the X-means algorithm [39], and the G-means algorithm [40]. Other popular clustering algorithms include hierarchical clustering (considered in fact a complete line of research with several algorithms available) [41], the Fuzzy C-Means algorithm [42], support vector clustering [43], decision trees clustering [44], and evolutionary clustering [45].

Clustering is one of the most popular research areas in data mining, so the reader should take the list as a small representative sample of the methods in the field. Please see Chapter 1 for more details on clustering algorithms or follow the avenues market by the revised methods.

2.3 Data Mining Applications in Power Systems

After a brief introduction to data mining and some of the most popular algorithms used in the field, we are ready to present some applications or "tasks" of these techniques in power systems. Since the KDD process has been widely applied in such systems, and due to the complexity of the current multidisciplinary environment (considering not only a wide variety of actors with different objectives but also an extensive range of technologies and engineering fields interconnected), we will limit to provide insights to some of the most common areas of applications, again preaching to the curious reader to go beyond the applications shown here, in order to understand the importance of such an area of research fully. Hopefully, the information provided here will be a good starting point to realize the reach of data mining in power systems.

2.3.1 Profiling

With the smart grid era, the electrical grid has suffered several technological changes, including the addition of a two-way flow of power and information. This has been achieved by installing sensors and an AMI that provides to different stakeholders a vast amount of information from the demand side. If used properly and efficiently, the information available allows a better understanding of the electrical consumption of end-users and can help improve the operation of the energy grid and enhance the grid stability [1].

Profiling in power systems is one of the most successful data mining applications due to the diverse tasks and benefits that it can provide. It achieved significant attention due to the large amount of available data in this new paradigm of smart grids and the impossibility of analyzing these data from the perspective of every single user. Therefore, and for the sake of having a common definition, load profiling was defined in 2001 by the International Energy Agency as [46]: "*The study of the consumption habits of consumers to estimate the amount of power they use at various times of the day and for which they are billed. Load profiling is an alternative to precise metering*." Of course, this definition is somehow limited these days since other considerations have emerged. Examples of these new considerations include the growing number of users equipped with AMI, consumers that are able to generate their own energy (i.e. prosumers), the emergence of local electricity markets, management of different time intervals and end users of different sizes, among others [47]. For the sake of coming into a common ground, we will refer to profiling as an electricity consumption pattern (either of consumption, generation, or net load) of end-users (independent of their size, e.g. residential, industrial, etc.) over a given period of time. In a more detailed and systematic view, load profiling can be divided into different stages, starting from load data preparation, applying clustering techniques to form groups according to their characteristic load pattern (see for instance Figure 2.1), evaluating such groups to come with some customer segmentation, to finally use the resulting profiles in a wide variety of scopes and applications [48, 49].

The different scopes of profiling applications have been evolving along with the available data, technological infrastructure, and new participants and stakeholders. For instance, profiling can play a key role in the settlement process by identifying the difference between the energy supplied and the actual energy used by end customers or determining such differences in the balancing process (i.e. for a balance responsible entity) [47]. These activities become relevant in networks in which not all end users are equipped with adequate AMI; therefore, classification of the data with alternative solutions is needed. Another example of profiling applications is tariff definition, an activity in which the profiles are used to formulate tariff structures applicable to the defined macro-clusters according to common patterns [50, 51]. Other scopes of applications include forecasting [52], demand response [1], or aggregated load modeling [53].

Going more in detail about applications that have emerged over the years, in Ref. [27], the authors proposed a new methodology (notice that some years have passed since this work was published) to create classes defined in the billing information space. The methodology used billing data from low voltage distribution clients (usually available in databases of distribution companies) and sample load diagrams only available at the time after a dedicated measurement campaign (a situation that might change with the adoption of AMI). In this methodology, the problem was modeled as an optimization problem with the objective of improving given criteria as a function of the obtained clusters. A similar methodology for load pattern shape clustering was later proposed in

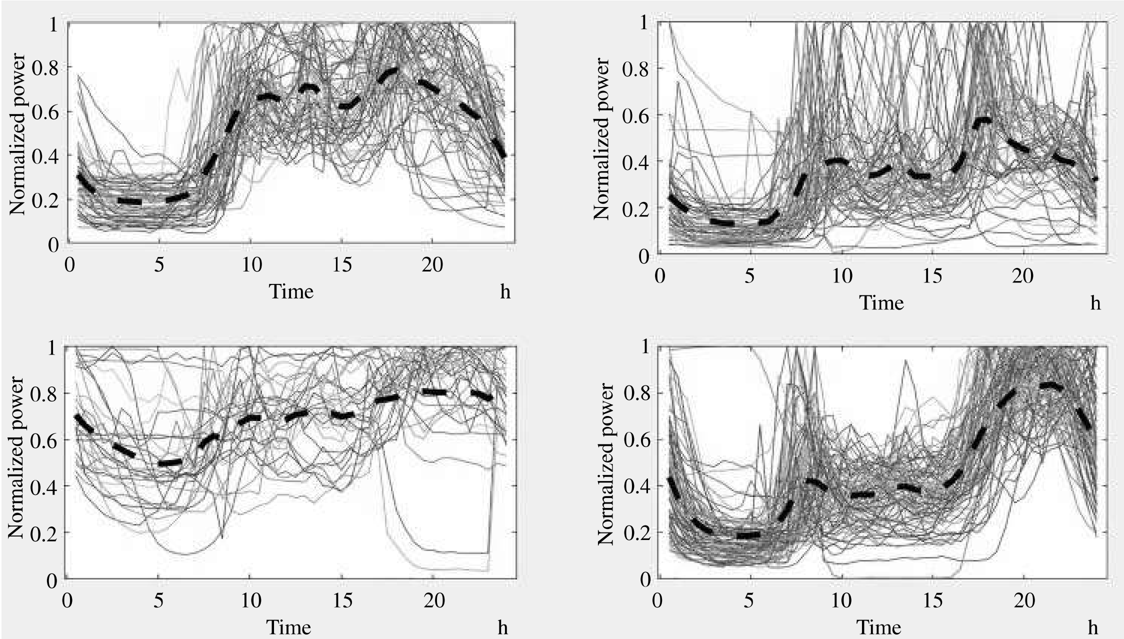

Figure 2.1 An example of load pattern shape clustering. Using k-means algorithm, the daily consumption of 250 end-users is grouped into four clusters (i.e. $k = 4$). The dotted line in each figure shows the resulting characteristic profile (i.e. the centroid of the cluster). Source: Adapted from Lezama et al. [48].

Refs. [48, 51]. In those works, new clustering algorithms were proposed to optimize a given "metric" or criteria rather than using existing clustering algorithms. These works were motivated by previous studies in which the characterization of the electrical patterns was done using various unsupervised and supervised learning techniques, namely k-means, fuzzy k-means, self-organizing maps, hierarchical clustering, etc. [54, 55]. The idea was that, despite the effectiveness of such available algorithms, the electrical load patterns have some unique characteristics (for instance, number of features, overlapping of patterns, etc.) that make their analysis different from the more general pattern recognition algorithms. Inspired by this idea, and following the trends in algorithmic design and technology, new proposals of novel algorithms, such as weighted Fuzzy C-Means [56], enhanced symbolic aggregate approximation [57], online (k-means) adaptive clustering [58], concurrent k-means and spectral clustering [59], just to name a few, have been proposed for customer characterization. It is also important to highlight the presence of the k-means algorithm in many of the studies that have clustering in their frameworks. Also, notice that while the main application of profiling was "load profiling" for customer characterization [28], this trend has changed over the years. With consideration of more sophisticated AMI, and the emergence of new actors in the energy chain such as prosumers, in current days profiling is not limited to load characterization [47]. Other applications of profiling can be, for instance, the determination of heating patterns in buildings [60], retail real pricing determination [61], and user classification considering electric vehicles [62].

Finally, we have summarized some of the mentioned applications in Table 2.1. The reader, however, is advised into the discovery of new techniques emerging in the context of power systems, taking into consideration that such systems are still experiencing vertiginous changes, evolving toward a complex socioeconomic environment in which synergy from different energy sources and stakeholders are not as clear and evident as they were in the past. Overall, load profiling will remain a crucial application of data mining in this new context, supporting the pursuit of an integrated solution that benefits energy efficiency and sustainability.

Table 2.1 Selected load profiling tasks and algorithms in power systems.

Year	Task	Algorithm	References
2012	Load profiling based on billing data	Clustering optimization	[27]
2012	Heating load patterns in buildings	k-means clustering	[60]
2016	Electrical load pattern shape clustering	Evolutionary clustering	[51]
2016	Clustering of smart grid customers	Weighted Fuzzy C-Means	[57]
2017	Energy patterns of office buildings	Fuzzy C-Means clustering	[28]
2018	Retail real-time pricing determination	Ensemble clustering	[61]
2018	Characterization of energy patterns	Enhanced sax	[57]
2019	Profiling of electricity and district heating	Online (k-means) adaptive clustering	[58]
2020	Household electricity consumption patterns	Concurrent k-means and spectral clustering	[59]
2021	User classification considering electric vehicles	k-means	[62]

2.3.2 Forecasting

Similar to "load" profiling, the application of forecasting in PES has become broader, including not only the forecasting of renewable generation (e.g. photovoltaic or wind-based generation) but also the forecast of load demand, market prices, or even electric vehicles trips based on human behavior [63–66]. The new challenges emerging in this domain are linked again to a variety of causes, passing from the need in the adoption of variable and uncertain renewable generation due to environmental issues (a need that is supported by governments worldwide) [52], the adoption of new technologies that allow access to a major amount of data from consumers [67], the deregulation of the power energy industry giving birth to new market structures and roles [65], just to name a few.

Moreover, forecasting techniques are usually inserted in more complex methodologies, such as a block of an energy management system. Therefore, the capability to deal with uncertainty directly impacts the risk and benefits that companies can assume. Another example of this interconnection of tasks can be a retailer's participation in the electricity market. In such a case, a comprehensive and accurate forecast of load demand, power generation, and electricity prices will provide a competitive advantage for the retailer, resulting in better prices for the end-user [68].

Despite the so-far efforts in the area, it is understandable that the present and future power grid requires more accurate forecasting models and tools to achieve optimal coordination between the different distributed resources. Thus, the research in such an area is vast, also supported by the appearance of novel algorithms belonging to the realm of artificial intelligence (see Figure 2.2). One of the most common forecasting applications is, in fact, load forecasting. In Ref. [70], a deep forest regression algorithm is used for short-term load forecasting (STLF) to mitigate the effect of hyper-parameters affecting deep learning algorithms. The authors recognize a large number of studies using ANN for load forecasting but also identify some limitations of ANN-based approaches related, for instance, to the long time required for model training, the configuration of many hyper-parameters, and the high probability of getting stuck in local optimum resulting

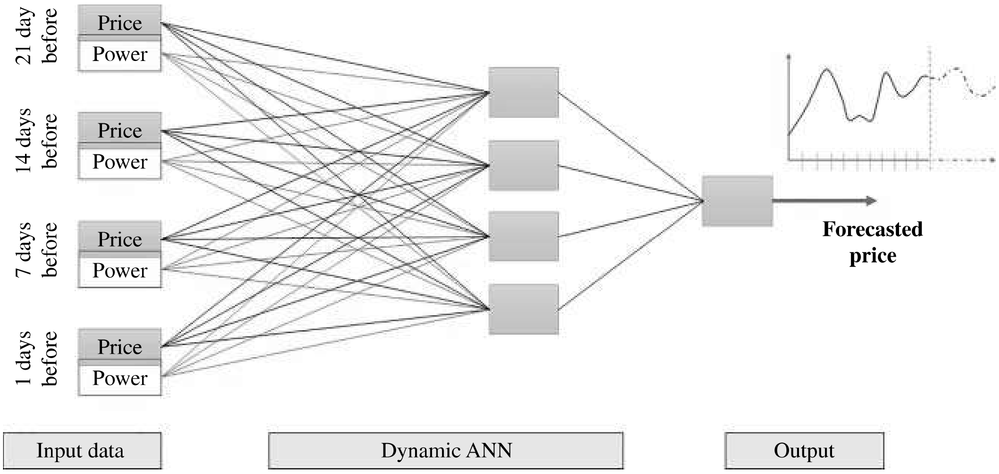

Figure 2.2 An example of an ANN that receives as inputs market prices and energy negotiated in the market (from one day, one week, two weeks, and three weeks before to the desired forecasted day). The ANN has four nodes in the intermediate layer and one output – the forecasted market price. Source: Adapted from Pinto et al. [69].

in high forecasting errors. While new techniques based on deep neural networks solve some of the aforementioned issues, the training speed is still low for short-term forecasts, and the need for parameter configuration is not avoided. More recently, in Ref. [71], a hybrid method combining machine learning techniques and evolutionary computation, named intelligent machine learning with an evolutionary algorithm-based short-term load forecasting (IMLEA-STLF) model, was proposed. This comprehensive approach makes use of other popular techniques in machine learning, such as wavelet transform for the decomposition of time series, oppositional artificial fish swarm optimization algorithm for feature selection, water wave optimization with Elman neural networks model for the prediction process, and finally inverse wavelet transform to obtain the resulting forecast. Results of such a complex method were satisfactory, beating other well-known algorithms commonly used for forecast, namely ARIMA, SVR, ARIMA-NN, and KNN. In the same line, [72] proposes a multi-temporal–spatial-scale method for STLF, using convolutional networks. The approach is compared with 22 forecasting algorithms (the reader can appreciate the vast number of approaches existing out there), demonstrating its capabilities to provide accurate results for the task.

Another popular task in which forecast is applied is the prediction of renewable generation. Renewables represent a sustainable and low-cost form of electricity, but their variability dependent on weather conditions (impossible to fully predict) pose some challenges in their adoption. It is interesting to find some initiatives such as the "global energy forecasting competition" [73], attracting hundreds of participants willing to test their algorithms aiming at achieving a more accurate forecast. From the different sources of renewable energy, wind and solar power are the most popular, and several works can be found in the literature in this regard. For instance, in Ref. [74], an ANN is used to forecast day-ahead wind power and locational marginal prices information. The consideration of wind power and locational marginal prices information, together with the management of energy storage systems, proved to be financially profitable for a wind farm. In Ref. [75], the authors assessed the effectiveness of an SVR model for solar power forecasting of a PV plant. A similar study was conducted in [76], comparing several forecasting algorithms such as KNN, ANN, SVR, etc. to forecast the power generation of a whole year of PV production of 32 plants with a wide variety of sizes and technologies. Other applications of data mining for forecast include the use of deep ensemble learning for probabilistic load forecast [77], the use of ANN, SVM, and radial basis

Table 2.2 Selected forecasting tasks and algorithms in power systems.

Year	Task	Algorithm	References
2015	Wind power forecast	ANN	[54]
2016	Prediction of net demand	Levenberg–Marquardt learning algorithm	[63]
2017	Solar power forecast	SVR	[75]
2018	Photovoltaic generation	KNN; ANN; SVR; and Quantile random forest	[76]
2019	Probabilistic load forecast	Deep ensemble learning	[77]
2019	Price and load forecast	ANN, SVM, and radial basis function neural network	[78]
2020	STLF	Deep forest regression	[70]
2021	STLF	Machine learning and evolutionary algorithms	[71]
2021	STLF	Convolutional network	[72]
2022	Renewable energy forecast	Normalizing flows	[79]

function neural network to provide a multi-block engine for load and price prediction [78], or the use of a novel probabilistic deep learning technique, normalizing flows, to compute accurate PV, wind, and load forecasting [79].

Finally, we summarize the analyzed applications in Table 2.2. Once more, the reader is encouraged to take this as an initial point in search of novel applications in the area, remembering that despite the recent advances, the perfect forecast in any of the domains (load, price, renewable generation, etc.) is still an open issue. While 100% of forecast accuracy is impossible to guarantee, many of the management systems, and why not to say the entire operation and interconnections of the energy grid in the smart grid era, depend on a large degree on the effectiveness of forecasting systems, so that designing better tools always gives an added value to the system as a whole.

2.3.3 Fault Detection and Diagnosis

As one of the most complex systems created by humans, the electrical grid suffers from many undesired variations in voltage and current due to a diverse number of factors such as impulses, notches, glitches, harmonic distortions, and of course, system faults. Moreover, with the emergence of new technologies, supported by the smart grid era, and the penetration of distributed generation (with a large load of uncontrollable renewables into the mix), concerns regarding power quality have grown recently [80].

System faults can be caused by different reasons and factors in the current era. For instance, up to 85% of system faults can be caused by issues occurring in transmission lines [29]. Other types of faults are also related to the penetration of new devices with a high voltage variability into the network. For instance, in 2016, the global installed capacity of PV generation worldwide was reported to be around 310 GW [81]. Therefore, with more penetration of these devices, fault detection and diagnosis (FDD) techniques of PV arrays became a relevant issue to tackle [29]. The same reasoning can be applied to almost all electrical grid components, making FDD a key application of data mining and power engineering in general.

Artificial intelligence and data mining (our topic of interest in this section) have also contributed with tools and techniques for fault detection and diagnosis. For instance, in buildings, a sector that accounts for approximately 39% of energy consumption in the United States and 40% in the European Union, automatic FDD is crucial to reduce service cost monitoring, equipment downtime, or energy penalties due to malfunction of a wide variety of devices, namely economizers, chillers, air handling units, heating, ventilation, and air conditioning (HVAC) systems, among others [82]. As it can be seen, FDD is a key application of data mining and artificial intelligence, as these two fields can provide effective tools for automated analysis to detect, locate, and diagnose faults of diverse types [83].

Now, analyzing in more detail some tasks and algorithms applied to this problem, we can cite, for instance, the work from Ref. [84], in which decision trees and SVM were used to detect and locate in real-time electricity theft in different levels of the power chain, from transmission to distribution. The application was tested showing that the number of false positives can be reduced to a large extent. Another task in which data mining can be used in this context is the detection of high impedance faults in distribution networks. The works of Refs. [85, 86] are a good example of these applications, using SVM, decision trees, and minimum description length-based approaches for high impedance faults detection. In a similar line of research, [87] proposes the use of Aprion, an algorithm for mining frequent item sets of Boolean association rules, with information collected in a fault search table for fault location. Switching a bit in the tasks, in Ref. [88], a stacked sparse autoencoder is proposed for line trip fault diagnosis. The authors claim that the stacked sparse autoencoder has a deep learning structure, using principal component analysis and SVM, that is more effective due to its network structure and layer-wise training mechanism. In a more specific work, Qiqi et al. [89] focuses on the detection of branch lines in 10 kV distribution networks, using principal component analysis (PCA) and the frequent pattern-growth algorithm to determine whether or not the line disconnection has occurred. It can be appreciated that the association of rules mining plays a key role in the fault location and diagnosis tasks, with application that goes from the location of line disconnections, but can be extended to, for instance, the work in Ref. [90], in which association of rules is used for the fault detection of building HVAC systems, or [91], in which rule mining is used to diagnose the outage root cause.

Finally, we summarize some selected publications working on FDD in Table 2.3. The table includes tasks and algorithms used in the context, but the reader is encouraged to take this just as a representative sample of the application. Also, while most of the analyzed works focused on distribution networks (mainly in the detection of faults in lines), there are many other tasks to be explored depending on the context of interest, such as the automatic fault detection of chillers in buildings [92] or the fault detection of single phase-earth feeders [93], just to name a few others. With this summary of the application, we hope to awaken the reader's curiosity, but at the same time, to make the reader realize about the importance and reach that FDD plays in the power system field.

2.3.4 Other Applications

As the reader can guess by now, the applications summarized and grouped by the authors in profiling, forecasting, and fault detection and diagnosis, are just a representative part of the reaches of data mining in power energy systems. This section will help the reader go beyond the categories mentioned above and appreciate other important applications in which studies and new approaches appear day by day.

For instance, security assessment of system stability is one of the more explored and studied topics in power systems, mainly because it is a useful task for many other applications such as day-ahead scheduling, real-time operation, and long-term planning [94]. While traditional methods for stability assessment rely on time-domain simulation requiring huge computational resources,

Table 2.3 Selected FDD tasks and algorithms in power systems.

Year	Task	Algorithm	References
2016	Detection and location electricity theft	Decision trees and SVM	[84]
2018	High impedance faults detection in distribution networks	SVM and decision trees	[86]
2018	Power system line trip fault diagnosis	PCA and SVM	[88]
2019	High impedance faults detection in distribution networks	Minimum description length-based discretization	[23]
2019	FDD of line ungrounded disconnection	PCA and frequent pattern-growth algorithm	[89]
2019	FDD of HVAC systems	Association rule comparison-based approach	[90]
2020	Smart FDD of outage root cause	Association rule-based method	[91]
2020	FDD in distribution networks	Aprior mining algorithm	[87]
2020	Automatic FDD for chillers	Generative adversarial network	[92]
2021	Single-phase earth fault feeder detection	Fuzzy C-Means, fuzzy measure fusion criterion	[93]

data mining opens the possibility of exploring data-driven approaches, using the measurements in the network and artificial intelligence to overcome this situation. Since early year, till now, the use of pattern recognition and data mining has opened a world of possibilities for security assessment [95–98]. Other examples of these alternative applications are the use of SVM for power system stability [99], or transient stability assessment [100]. Clustering techniques have also been used for operational performance in buildings [101], or for economic dispatch [102], and we can even cite applications of machine learning for the design of autonomous brokers participating in electricity markets [103].

To be coherent with the structure of this chapter, we summarize selected task applications in power systems in Table 2.4. It is perhaps pertinent to recall that, as the reader has already noticed, the line dividing the tasks and application is a bit gray since all the devices and systems coexist in the current electricity system. Therefore, it is interesting to do an exercise trying to position the diverse approaches and methodologies exposed here in the big panorama of power energy systems. In fact, power system visualization [106, 107], another field of application in the field, uses a wide variety of data mining algorithms to simplify models and give a better presentation and interpretation of the huge amount of available data nowadays.

2.4 Discussion and Final Remarks

We have analyzed in this chapter several applications of data mining in power systems and highlighted the crucial role that this data science plays in the future of the electrical grid. Accordantly, we put together some of the most representative algorithms in the area, realizing that some applications require even combinations of approaches that respond to a bigger framework. For instance,

Table 2.4 Selected applications in diverse topics on power systems.

Year	Task	Algorithm	References
1989	Security assessment	Adaptive pattern recognition, ANN	[96]
2009	Security assessment	Decision trees	[95]
2014	Building operational performance	Clustering analysis	[101]
2018	Power system stability	Modified SVM	[99]
2018	Protection of distribution networks	Decision trees and adaptive NN	[104]
2018	Power quality problems	Decision trees algorithms in WEKA	[105]
2018	Transient stability assessment	Convolutional SVM	[100]
2019	Autonomous broker for energy markets	Reinforcement learning	[103]
2021	Security assessment	Random bits forest	[97]
2021	Stability and security management	Machine learning and classification models	[98]
2022	Economic load dispatch	Hierarchical clustering	[102]

imagine an energy management system in which a block is in charge of forecasting renewable generation, another block in charge of user profiling, another block in charge of market participation, another block in charge of network reliability and assessment, and so on.

Since the information provided in this chapter is extensive, and we recognize that it could be easy for the reader to get lost in the concepts provided, we have put graphically, in Figure 2.3, the taxonomy resulting from this study. In here, we can first appreciate in the highest level of the KDD process, with its five steps, being data mining a key step of the process. After that, in the second level from top to button, we divided data mining into two categories: supervised and unsupervised data mining algorithms. Supervised (learning) is referred to algorithms that require "a teacher" meaning that these algorithms determine outputs variables for a given set of input or predictor variables.

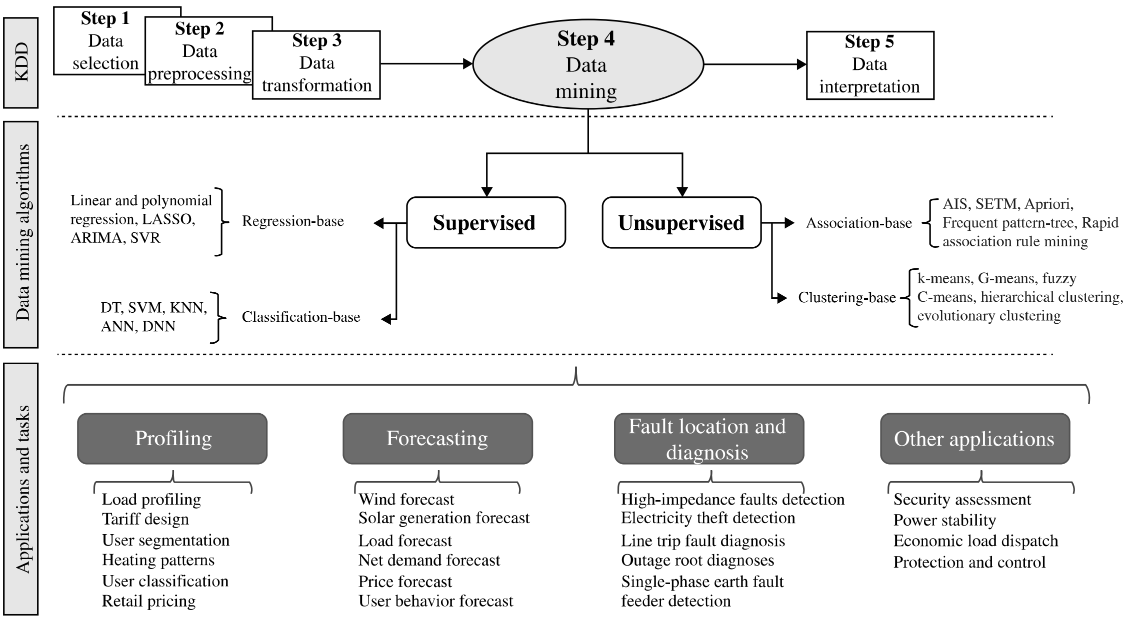

Figure 2.3 Taxonomy of data mining applications in PES.

The two main groups of algorithms belonging to supervised data mining are regression-based methods (e.g. linear regression, LASSO, SVR) and classification-based methods (e.g. decision trees, SVM, KNN). On the other hand, unsupervised (learning) is referred to algorithms that do not require "a teacher" meaning that these algorithms find (or try to find) the underlying structure of input data without knowing the corresponding output variables. Unsupervised data mining algorithms can be classified into association rule-based approaches (e.g. Apriori algorithm, and several types of ARM), and clustering-based approaches (e.g. k-means, G-means, hierarchical clustering). After putting in order some of the (several) available algorithms for data mining, we proceeded to a third level to the application of these methods in power systems. Again, due to the extensive available literature, we came up with some categories that we believe are still relevant in the design and management of power systems, namely profiling, forecasting, and fault detection and diagnosis. For each of these categories, we provided a summary of selected applications, emphasizing the algorithms used and the purpose of the task. We also included in Section 2.3.4, other applications out of these categories (e.g. security assessment, network reliability, etc.) for the reader to have a glimpse into the broader picture of the role of data mining in power systems.

To conclude, we would like to remind the reader about the important role that PES play in the world. Despite the technological developments in the area, improper operation, control, and equipment faults represent a serious energy waste that needs to be handled to pursue a sustainable energy road. The recent advances in knowledge discovery, and the current trends of automation systems, putting the "smart" in the electrical grid, have led to a massive amount of available data. Thus, the use and extraction of knowledge from available data become fundamental. As was previously mentioned, data mining plays a crucial role in the KDD process and is a very promising tool to extract and synthesize information coming from the complex socio-economical system that power systems represent. We certainly hope that this chapter has provided the reader the means and information to pursue this line of research and devise novel applications resulting from the existing ones, an avenue we believe is far from reaching its maximum potential.

References

1. Yi Wang, Qixin Chen, Chongqing Kang, Mingming Zhang, Ke Wang, and Yun Zhao. 2015. "Load profiling and its application to demand response: a review." *Tsinghua Science and Technology* 20 (2): 117–29. https://doi.org/10.1109/TST.2015.7085625.
2. Witten, I.H., Frank, E., and Hall, M.A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier https://doi.org/10.1016/C2009-0-19715-5.
3. Ibrahim, Muhammad Sohail, Wei Dong, and Qiang Yang. 2020. "Machine learning driven smart electric power systems: current trends and new perspectives." *Applied Energy* 272 115237. https://doi.org/10.1016/j.apenergy.2020.115237.
4. Overbye, T.J. and Weber, J.D. (2000). Visualization of power system data. In: *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, vol. 1, 7. IEEE https://doi.org/10.1109/HICSS.2000.926744.
5. Zhao, Yang, Chaobo Zhang, Yiwen Zhang, Zihao Wang, and Junyang Li. 2020. "A review of data mining technologies in building energy systems: load prediction, pattern identification, fault detection and diagnosis." *Energy and Built Environment* 1 (2): 149–64. https://doi.org/10.1016/j.enbenv.2019.11.003.
6. Saravanan, R., and Pothula Sujatha. 2018. "A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification." In *2018 Second*

International Conference on Intelligent Computing and Control Systems (ICICCS), 945–49. IEEE. https://doi.org/10.1109/ICCONS.2018.8663155.

7 Amasyali, Kadir, and Nora M. El-Gohary. 2018. "A review of data-driven building energy consumption prediction studies." *Renewable and Sustainable Energy Reviews* 81 1192–1205. https://doi.org/10.1016/j.rser.2017.04.095.

8 Zhao, Hai-Xiang, and Frédéric Magoulès. 2012. "A review on the prediction of building energy consumption." *Renewable and Sustainable Energy Reviews* 16 (6): 3586–92. https://doi.org/10.1016/j.rser.2012.02.049.

9 Granados-Lieberman, D., R.J. Romero-Troncoso, R.A. Osornio-Rios, A. Garcia-Perez, and E. Cabal-Yepez. 2011. "Techniques and methodologies for power quality analysis and disturbances classification in power systems: a review." *IET Generation Transmission and Distribution* 5 (4): 519. https://doi.org/10.1049/iet-gtd.2010.0466.

10 García-Martos, C. and Conejo, A.J. (2013). Price forecasting techniques in power systems. In: *Wiley Encyclopedia of Electrical and Electronics Engineering*. Hoboken, NJ: Wiley https://doi.org/10.1002/047134608X.W8188.

11 Dasgupta, Abhijit, Yan V. Sun, Inke R. König, Joan E. Bailey-Wilson, and James D. Malley. 2011. "Brief review of regression-based and machine learning methods in genetic epidemiology: the genetic analysis workshop 17 experience." *Genetic Epidemiology* 35 (S1): S5–11. https://doi.org/10.1002/gepi.20642.

12 Ghiassi, M., H. Saidane, and D.K. Zimbra. 2005. "A dynamic artificial neural network model for forecasting time series events." *International Journal of Forecasting* 21 (2): 341–62. https://doi.org/10.1016/j.ijforecast.2004.10.008.

13 Murata, N., S. Yoshizawa, and S. Amari. 1994. "Network information criterion-determining the number of hidden units for an artificial neural network model." *IEEE Transactions on Neural Networks* 5 (6): 865–72. https://doi.org/10.1109/72.329683.

14 Tibshirani, Robert. 1996. "Regression shrinkage and selection via the Lasso." *Journal of the Royal Statistical Society: Series B: Methodological* 58 (1): 267–88. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

15 Box, G. E. P., and David A. Pierce. 1970. "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models." *Journal of the American Statistical Association* 65 (332): 1509–26. https://doi.org/10.1080/01621459.1970.10481180.

16 Smola, Alex J., and Bernhard Schölkopf. 2004. "A tutorial on support vector regression." *Statistics and Computing* 14 (3): 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.

17 Miikkulainen, R., Liang, J., Meyerson, E. et al. (2019). Evolving deep neural networks. In: *Artificial Intelligence in the Age of Neural Networks and Brain Computing* (ed. Kozma, R., Alippi, C., Choe, Y., and Morabito, F.C.), 293–312. Academic Press, Elsevier. https://doi.org/10.1016/B978-0-12-815480-9.00015-3.

18 Adamowski, Jan, Fung Chan, H. Prasher, S. O. Ozga-Zielinski, B. and Sliusarieva, A. 2012. "Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada." *Water Resources Research* 48 (1). https://doi.org/10.1029/2010WR009945.

19 Boussaada, Zina, Octavian Curea, Ahmed Remaci, Haritza Camblong, and Najiba Mrabet Bellaaj. 2018. "A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of the daily direct solar radiation." *Energies* 11 (3): 620. https://doi.org/10.3390/en11030620.

20 Marquardt, Donald W., and Ronald D. Snee. 1975. "Ridge regression in practice." *The American Statistician* 29 (1): 3–20. https://doi.org/10.1080/00031305.1975.10479105.

21 Matloff, N. (2017). *Statistical Regression and Classification: From Linear Models to Machine Learning*. CRC Press.

22 Hosmer, D.W. Jr., Lemeshow, S., and Sturdivant, R.X. (2013). *Applied Logistic Regression*. Wiley.

23 Noble, William S. 2006. “What is a support vector machine?” *Nature Biotechnology* 24 (12): 1565–67. https://doi.org/10.1038/nbt1206-1565.

24 Quinlan, J.R. 1990. “Decision trees and decision-making.” *IEEE Transactions on Systems, Man, and Cybernetics* 20 (2): 339–46. https://doi.org/10.1109/21.52545.

25 Cunningham, Pádraig, and Sarah Jane Delany. 2022. “K-nearest neighbour classifiers – a tutorial.” *ACM Computing Surveys* 54 (6): 1–25. https://doi.org/10.1145/3459665.

26 Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. “Unsupervised learning.” In *The Elements of Statistical Learning*, 1–101. New York, NY: Springer. https://doi.org/10.1007/b94608_14.

27 Fidalgo, José Nuno, Manuel António Matos, and Luís Ribeiro. 2012. “A new clustering algorithm for load profiling based on billing data.” *Electric Power Systems Research* 82 (1): 27–33. https://doi.org/10.1016/j.epsr.2011.08.016.

28 Qin, J., Zhang, J. 2017. “Sampling for building energy consumption with fuzzy theory.” *Energy and Buildings* 156 78–84. https://doi.org/10.1016/j.enbuild.2017.09.047.

29 Mellit, A., G.M. Tina, and S.A. Kalogirou. 2018. “Fault detection and diagnosis methods for photovoltaic systems: a review.” *Renewable and Sustainable Energy Reviews* 91 1–17. https://doi.org/10.1016/j.rser.2018.03.062.

30 Agrawal, Rakesh, Tomasz Imieliński, and Arun Swami. 1993. “Mining association rules between sets of items in large databases.” In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data – SIGMOD ’93*, 207–16. New York, NY: ACM Press. https://doi.org/10.1145/170035.170072.

31 Agrawal, R., T. Imieliński, and A. Swami. 1993. “Database mining: a performance perspective.” *IEEE Transactions on Knowledge and Data Engineering* 5 (6): 914–25. https://doi.org/10.1109/69.250074.

32 Houtsma, M. and Swami, A. (1995). Set-oriented mining for association rules in relational databases. In: *Proceedings of the Eleventh International Conference on Data Engineering*, 25–33. IEEE https://doi.org/10.1109/ICDE.1995.380413.

33 Agrawal, R. and Ramakrishnan, S. (1994). Fast algorithms for mining association rules. In: *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, 487–499. Morgan Kaufmann.

34 Rodríguez-González, A.Y., Lezama, F., Iglesias-Alvarez C.A., et al. 2018. “Closed frequent similar pattern mining: reducing the number of frequent similar patterns without information loss.” *Expert Systems with Applications* 96 271–83. https://doi.org/10.1016/j.eswa.2017.12.018.

35 Zhao, Q. and Bhowmick, S.S. (2003). Association rule mining: a survey. Technical report 135, Nanyang Technological University, Singapore.

36 Likas, Aristidis, Nikos Vlassis, and Jakob J. Verbeek. 2003. “The global K-means clustering algorithm.” *Pattern Recognition* 36 (2): 451–61. https://doi.org/10.1016/S0031-3203(02)00060-2.

37 Peña, J.M, J.A Lozano, and P Larrañaga. 1999. “An empirical comparison of four initialization methods for the K-means algorithm.” *Pattern Recognition Letters* 20 (10): 1027–40. https://doi.org/10.1016/S0167-8655(99)00069-0.

38 Wu, Kuo-Lung, and Miin-Shen Yang. 2007. “Mean shift-based clustering.” *Pattern Recognition* 40 (11): 3035–52. https://doi.org/10.1016/j.patcog.2007.02.006.

39 Pelleg, D. and Moore, A.W. (2000). X-means: extending K-means with efficient estimation of the number of clusters. In *Proceedings of the Seventeenth International Conference on Machine Learning (ICML '00)*, 727–734. San Francisco, CA: Morgan Kaufmann Publishers Inc.

40 Hamerly, G. and Elkan, C. (2003). Learning the k in K-means. In: *Proceedings of the 16th International Conference on Neural Information Processing Systems*, 281–288. MIT Press.

41 Murtagh, Fionn, and Pedro Contreras. 2011. "Algorithms for hierarchical clustering: an overview." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2 (1): 86–97. https://doi.org/10.1002/widm.53.

42 Dunn, J. C. 1973. "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters." *Journal of Cybernetics* 3 (3): 32–57. https://doi.org/10.1080/01969727308546046.

43 Ben-Hur, A., Horn, D., Siegelmann, H.T., and Vapnik, V. (2001). Support vector clustering. *Journal of Machine Learning Research* 2: 125–137.

44 Liu, B., Xia, Y., and Philip, S.Y. (2000). Clustering through decision tree construction. In: *Proceedings of the Ninth International Conference on Information and Knowledge Management*, 20–29. Association for Computing Machinery.

45 Handl, Julia, and Joshua Knowles. 2007. "An evolutionary approach to multiobjective clustering." *IEEE Transactions on Evolutionary Computation* 11 (1): 56–76. https://doi.org/10.1109/TEVC.2006.877146.

46 IEA – International Energy Agency (2001). *Competition in Electricity Markets*. Organisation for Economic Co-operation and Development (OECD).

47 Chicco, Gianfranco, and Andrea Mazza. 2021. "Load profiling revisited: prosumer profiling for local energy markets." In *Local Electricity Markets*, 215–42. Elsevier. https://doi.org/10.1016/B978-0-12-820074-2.00004-6.

48 Lezama, Fernando, Ansel Y. Rodriguez-Gonzalez, and de Cote, E.M. 2016. "Load pattern clustering using differential evolution with pareto tournament." In *2016 IEEE Congress on Evolutionary Computation (CEC)*, 241–48. IEEE. https://doi.org/10.1109/CEC.2016.7743801.

49 Chicco, G., R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader. 2004. "Load pattern-based classification of electricity customers." *IEEE Transactions on Power Systems* 19 (2): 1232–39. https://doi.org/10.1109/TPWRS.2004.826810.

50 Chicco, G., R. Napoli, P. Postolache, M. Scutariu, and C. Toader. 2003. "Customer characterization options for improving the tariff offer." *IEEE Transactions on Power Systems* 18 (1): 381–87. https://doi.org/10.1109/TPWRS.2002.807085.

51 Lezama, F., A.Y. Rodríguez, E.M. de Cote, and L.E. Sucar. 2016. Electrical load pattern shape clustering using ant colony optimization. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* Vol. 9597. https://doi.org/10.1007/978-3-319-31204-0_32.

52 Ahmed, Adil, and Muhammad Khalid. 2019. "A review on the selected applications of forecasting models in renewable power systems." *Renewable and Sustainable Energy Reviews* 100 9–21. https://doi.org/10.1016/j.rser.2018.09.046.

53 Lu, Ning. 2012. "An evaluation of the HVAC load potential for providing load balancing service." *IEEE Transactions on Smart Grid* 3 (3): 1263–70. https://doi.org/10.1109/TSG.2012.2183649.

54 Figueiredo, V., Rodrigues, F., Vale, Z., and Gouveia, J.B. (2005). An electric energy consumer characterization framework based on data mining techniques. *IEEE Transactions on Power Systems* 20 (2): 596–602.

55 Chicco, Gianfranco. 2012. "Overview and performance assessment of the clustering methods for electrical load pattern grouping." *Energy* 42 (1): 68–80. https://doi.org/10.1016/j.energy.2011.12.031.

56 Lu, Yun, Tiankui Zhang, and Zhimin Zeng. 2016. "Adaptive weighted fuzzy clustering algorithm for load profiling of smart grid customers." In *2016 IEEE/CIC International Conference on Communications in China (ICCC)*, 1–6. IEEE. https://doi.org/10.1109/ICCChina.2016.7636874.

57 Capozzoli, Alfonso, Marco Savino Piscitelli, Silvio Brandi, Daniele Grassi, and Gianfranco Chicco. 2018. "Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings." *Energy* 157 336–52. https://doi.org/10.1016/j.energy.2018.05.127.

58 Ray, G. Le, and P. Pinson. 2019. "Online adaptive clustering algorithm for load profiling." *Sustainable Energy, Grids and Networks* 17 100181. https://doi.org/10.1016/j.segan.2018.100181.

59 Sun, Li, Kaile Zhou, and Shanlin Yang. 2020. "An ensemble clustering based framework for household load profiling and driven factors identification." *Sustainable Cities and Society* 53 101958. https://doi.org/10.1016/j.scs.2019.101958.

60 Nikolaou, Triantafyllia G., Dionysia S. Kolokotsa, George S. Stavrakakis, and Ioannis D. Skias. 2012. "On the application of clustering techniques for office Buildings' energy and thermal comfort classification." *IEEE Transactions on Smart Grid* 3 (4): 2196–2210. https://doi.org/10.1109/TSG.2012.2215059.

61 Joseph, Shibily, and Jasmin Erakkath Abdu. 2018. "Real-time retail price determination in smart grid from real-time load profiles." *International Transactions on Electrical Energy Systems* 28 (3): e2509. https://doi.org/10.1002/etep.2509.

62 Ahmed, Saeed, Zafar Ali Khan, Noor Gul, Junsu Kim, and Su Min Kim. 2021. "Machine learning-based clustering of load profiling to study the impact of electric vehicles on smart meter applications." *2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN)*, 444–47. IEEE. https://doi.org/10.1109/ICUFN49451.2021.9528396.

63 Abedinia, O., and Amjady, N. 2016. "Net demand prediction for power systems by a new neural network-based forecasting engine." *Complexity* 21 (S2): 296–308. https://doi.org/10.1002/cplx.21807.

64 Morsalin, Sayidul, Khizir Mahmud, and Graham Town. 2016. "Electric vehicle charge scheduling using an artificial neural network." In *2016 IEEE Innovative Smart Grid Technologies – Asia (ISGT-Asia)*, 276–80. IEEE. https://doi.org/10.1109/ISGT-Asia.2016.7796398.

65 Negnevitsky, M., Mandal, P., and Srivastava, A.K. (2009). An overview of forecasting problems and techniques in power systems. In *2009 IEEE Power & Energy Society General Meeting*, 1–4. IEEE. https://doi.org/10.1109/PES.2009.5275480.

66 Rashid, M.H. (2018). AMI smart meter big data analytics for time series of electricity consumption. In: *In 2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE)*, 1771–1776. IEEE https://doi.org/10.1109/TrustCom/BigDataSE.2018.00267.

67 Quilumba, Franklin L., Wei-Jen Lee, Heng Huang, David Y. Wang, and Robert L. Szabados. 2015. "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities." *IEEE Transactions on Smart Grid* 6 (2): 911–18. https://doi.org/10.1109/TSG.2014.2364233.

68 Catalão, J.P. (ed.) (2017). *Electric Power Systems: Advanced Forecasting Techniques and Optimal Generation Scheduling*. CRC Press.

69 Pinto, T., Sousa, T.M., and Vale, Z. (2012). Dynamic artificial neural network for electricity market prices forecast. In: *2012 IEEE 16th International Conference on Intelligent Engineering Systems (INES)*, 311–316. IEEE.

70 Yin, Linfei, Zhixiang Sun, Fang Gao, and Hui Liu. 2020. "Deep forest regression for short-term load forecasting of power systems." *IEEE Access* 8: 49090–99. https://doi.org/10.1109/ACCESS.2020.2979686.

71 Mehedi, Ibrahim M., Hussain Bassi, Muhyaddin J. Rawa, Mohammed Ajour, Abdullah Abusorrah, Mahendiran T. Vellingiri, Zainal Salam, and Md. Pauzi Bin Abdullah. 2021. "Intelligent machine learning with evolutionary algorithm based short term load forecasting in power systems." *IEEE Access* 9: 100113–24. https://doi.org/10.1109/ACCESS.2021.3096918.

72 Yin, Linfei, and Jiaxing Xie. 2021. "Multi-temporal-spatial-scale temporal convolution network for short-term load forecasting of power systems." *Applied Energy* 283 116328. https://doi.org/10.1016/j.apenergy.2020.116328.

73 Hong, Tao, Pierre Pinson, and Shu Fan. 2014. "Global energy forecasting competition 2012." *International Journal of Forecasting* 30 (2): 357–63. https://doi.org/10.1016/j.ijforecast.2013.07.001.

74 Liu, Meng, Franklin L. Quilumba, and Wei-Jen Lee. 2015. "Dispatch scheduling for a wind farm with hybrid energy storage based on wind and LMP forecasting." *IEEE Transactions on Industry Applications* 51 (3): 1970–77. https://doi.org/10.1109/TIA.2014.2372043.

75 Antonanzas, J., D. Pozo-Vázquez, L.A. Fernandez-Jimenez, and F.J. Martinez-de-Pison. 2017. "The value of day-ahead forecasting for photovoltaics in the Spanish electricity market." *Solar Energy* 158 140–46. https://doi.org/10.1016/j.solener.2017.09.043.

76 Gigoni, Lorenzo, Alessandro Betti, Emanuele Crisostomi, Alessandro Franco, Mauro Tucci, Fabrizio Bizzarri, and Debora Mucci. 2018. "Day-ahead hourly forecasting of power generation from photovoltaic plants." *IEEE Transactions on Sustainable Energy* 9 (2): 831–42. https://doi.org/10.1109/TSTE.2017.2762435.

77 Yang, Yandong, Weijun Hong, and Shufang Li. 2019. "Deep ensemble learning based probabilistic load forecasting in smart grids." *Energy* 189 116324. https://doi.org/10.1016/j.energy.2019.116324.

78 Gao, Wei, Ayda Darvishan, Mohammad Toghani, Mohsen Mohammadi, Oveis Abedinia, and Noradin Ghadimi. 2019. "Different states of multi-block based forecast engine for price and load prediction." *International Journal of Electrical Power & Energy Systems* 104 423–35. https://doi.org/10.1016/j.ijepes.2018.07.014.

79 Dumas, Jonathan, Antoine Wehenkel, Damien Lanaspeze, Bertrand Cornélusse, and Antonio Sutera. 2022. "A deep generative model for probabilistic energy forecasting in power systems: normalizing flows." *Applied Energy* 305 117871. https://doi.org/10.1016/j.apenergy.2021.117871.

80 Aleem, Saad Abdul, Nauman Shahid, and Ijaz Haider Naqvi. 2015. "Methodologies in power systems fault detection and diagnosis." *Energy Systems* 6 (1): 85–108. https://doi.org/10.1007/s12667-014-0129-1.

81 Appiah, Albert Yaw, Xinghua Zhang, Ben Beklisi Kwame Ayawli, and Frimpong Kyeremeh. 2019. "Review and performance evaluation of photovoltaic array fault detection and diagnosis techniques." *International Journal of Photoenergy* 2019 1–19. https://doi.org/10.1155/2019/6953530.

82 Zhao, Yang, Tingting Li, Xuejun Zhang, and Chaobo Zhang. 2019. "Artificial intelligence-based fault detection and diagnosis methods for building energy systems: advantages, challenges and the future." *Renewable and Sustainable Energy Reviews* 109 85–101. https://doi.org/10.1016/j.rser.2019.04.021.

83 Furse, Cynthia M., Moussa Kafal, Reza Razzaghi, and Yong-June Shin. 2021. “Fault diagnosis for electrical systems and power networks: a review.” *IEEE Sensors Journal* 21 (2): 888–906. https://doi.org/10.1109/JSEN.2020.2987321.

84 Jindal, Anish, Amit Dua, Kuljeet Kaur, Mukesh Singh, Neeraj Kumar, and S. Mishra. 2016. “Decision tree and SVM-based data analytics for theft detection in smart grid.” *IEEE Transactions on Industrial Informatics* 12 (3): 1005–16. https://doi.org/10.1109/TII.2016.2543145.

85 Cui, Qiushi, Khalil El-Arroudi, and Yang Weng. 2019. “A feature selection method for high impedance fault detection.” *IEEE Transactions on Power Delivery* 34 (3): 1203–15. https://doi.org/10.1109/TPWRD.2019.2901634.

86 Mohammadnian, Youness, Turaj Amraee, and Alireza Soroudi. 2019. “Fault detection in distribution networks in presence of distributed generations using a data mining-driven wavelet transform.” *IET Smart Grid* 2 (2): 163–71. https://doi.org/10.1049/iet-stg.2018.0158.

87 Xue, Bing, Qing Chen, and Wudi Huang. 2020. “A fault diagnosis method of active distribution network based on fault search table and data mining technology.” *2020 IEEE Sustainable Power and Energy Conference (ISPEC)*, 2512–18. IEEE. https://doi.org/10.1109/iSPEC50848.2020.9350932.

88 Wang, Yixing, Meiqin Liu, Zhejing Bao, and Senlin Zhang. 2019. “Stacked sparse autoencoder with PCA and SVM for data-based line trip fault diagnosis in power systems.” *Neural Computing and Applications* 31 (10): 6719–31. https://doi.org/10.1007/s00521-018-3490-5.

89 Qiqi, Z., Siyi, L., and Yan, Z. 2019. “Power distribution network disconnection fault diagnosis using data mining method.” In *2019 IEEE Innovative Smart Grid Technologies – Asia (ISGT Asia)*, 407–11. IEEE. https://doi.org/10.1109/ISGT-Asia.2019.8881320.

90 Zhang, Chaobo, Xue Xue, Yang Zhao, Xuejun Zhang, and Tingting Li. 2019. “An improved association rule mining-based method for revealing operational problems of building heating, ventilation and air conditioning (HVAC) systems.” *Applied Energy* 253 113492. https://doi.org/10.1016/j.apenergy.2019.113492.

91 Dehbozorgi, Mohammad Reza, and Mohammad Rastegar. 2020. “Association rule mining application to diagnose smart distribution power system outage root cause.” In *2020 10th Smart Grid Conference (SGC)*, 1–6. IEEE. https://doi.org/10.1109/SGC52076.2020.9335746.

92 Yan, Ke, Adrian Chong, and Yuchang Mo. 2020. “Generative adversarial network for fault detection diagnosis of chillers.” *Building and Environment* 172 106698. https://doi.org/10.1016/j.buildenv.2020.106698.

93 Yu, Kun, Hao Zou, Xiangjun Zeng, Yueyu Li, Hao Li, Chao Zhuo, and Zhan Wang. 2021. “Faulty feeder detection of single phase-earth fault based on fuzzy measure fusion criterion for distribution networks.” *International Journal of Electrical Power & Energy Systems* 125 106459. https://doi.org/10.1016/j.ijepes.2020.106459.

94 You, Shutang, Yinfeng Zhao, Mirka Mandich, Yi Cui, Hongyu Li, Huangqing Xiao, Summer Fabus, et al. 2020. “A review on artificial intelligence for grid stability assessment.” In *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 1–6. IEEE. https://doi.org/10.1109/SmartGridComm47815.2020.9302990.

95 Diao, Ruisheng, Vijay Vittal, and Naim Logic. 2010. “Design of a real-time security assessment tool for situational awareness enhancement in modern power systems.” *IEEE Transactions on Power Systems* 25 (2): 957–65. https://doi.org/10.1109/TPWRS.2009.2035507.

96 Sobajic, D.J., and Y.-H. Pao. 1989. “Artificial neural-net based dynamic security assessment for electric power systems.” *IEEE Transactions on Power Systems* 4 (1): 220–28. https://doi.org/10.1109/59.32481.

97 Liu, Songkai, Lihuang Liu, Nan Yang, Dan Mao, Lei Zhang, Jiangzhou Cheng, Tianliang Xue, et al. 2021. “A data-driven approach for online dynamic security assessment with spatial-temporal dynamic visualization using random bits Forest.” *International Journal of Electrical Power & Energy Systems* 124 106316. https://doi.org/10.1016/j.ijepes.2020.106316.

98 Zhang, C., Yuan, Z., and Yan, P. (2021). Analysis of smart grid stability and security management based on data mining. *IOP Conference Series: Earth and Environmental Science* 651: 022049.

99 Hou, Kaiyuan, Guanghui Shao, Haiming Wang, Le Zheng, Qiang Zhang, Shuang Wu, and Wei Hu. 2018. “Research on practical power system stability analysis algorithm based on modified SVM.” *Protection and Control of Modern Power Systems* 3 (1): 11. https://doi.org/10.1186/s41601-018-0086-0.

100 Bashiri Mosavi, Alireza, Ali Amiri, and Seyed Hadi Hosseini. 2018. “A learning framework for size and type independent transient stability prediction of power system using twin convolutional support vector machine.” *IEEE Access* 6: 69937–47. https://doi.org/10.1109/ACCESS.2018.2880273.

101 Xiao, Fu, and Cheng Fan. 2014. “Data mining in building automation system for improving building operational performance.” *Energy and Buildings* 75: 109–18. https://doi.org/10.1016/j.enbuild.2014.02.005.

102 Dai, Bangwu, Fuli Wang, and Yuqing Chang. 2022. “Multi-objective economic load dispatch method based on data mining technology for large coal-fired power plants.” *Control Engineering Practice* 121 105018. https://doi.org/10.1016/j.conengprac.2021.105018.

103 Rodríguez González, A.Y., M. Palacios Alonso, F. Lezama, L. Rodríguez, E. Muñoz de Cote, E.F. Morales, L. Enrique Sucar, and D.D. Crockett. 2019. “A competitive and profitable multi-agent autonomous broker for energy markets.” *Sustainable Cities and Society* 49. https://doi.org/10.1016/j.scs.2019.101590.

104 Tang, Wen-Jun, and Hong-Tzer Yang. 2018. “Data mining and neural networks based self-adaptive protection strategies for distribution systems with DGs and FCLs.” *Energies* 11 (2): 426. https://doi.org/10.3390/en11020426.

105 Asha Kiranmai, S., and A. Jaya Laxmi. 2018. “Data mining for classification of power quality problems using WEKA and the effect of attributes on classification accuracy.” *Protection and Control of Modern Power Systems* 3 (1): 29. https://doi.org/10.1186/s41601-018-0103-3.

106 Talaat, M., Abdulaziz S. Alsayyari, Adel Alblawi, and A.Y. Hatata. 2020. “Hybrid-cloud-based data processing for power system monitoring in smart grids.” *Sustainable Cities and Society* 55 102049. https://doi.org/10.1016/j.scs.2020.102049.

107 Zhu, Jun, Eric Zhuang, Chavdar Ivanov, and Ziwen Yao. 2011. “A data-driven approach to interactive visualization of power systems.” *IEEE Transactions on Power Systems* 26 (4): 2539–46. https://doi.org/10.1109/TPWRS.2011.2119499.

3

Deep Learning in Intelligent Power and Energy Systems

Bruno Mota, Tiago Pinto, Zita Vale, and Carlos Ramos

GECAD Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development (GECAD), Polytechnic Institute of Porto (ISEP/IPP), Porto, Portugal

Acronyms

AE	autoencoder
AI	artificial intelligence
ANN	artificial neural network
CNN	convolution neural network
DBN	deep belief network
DDPG	deep deterministic policy gradient
DL	deep learning
DNN	deep neural network
DQN	deep Q network
DR	demand response
DRL	deep reinforcement learning
EMS	energy management system
FD	fault detection
GRU	gated recurrent unit
HVAC	heating, ventilation, and air conditioning
IoT	Internet-of-Things
LSTM	long short-term memory
MDP	Markov decision process
ML	machine learning
PES	power and energy systems
PV	photovoltaic
RBM	restricted Boltzmann machine
RNN	recurrent neural network
SVM	support vector machine

Intelligent Data Mining and Analysis in Power and Energy Systems: Models and Applications for Smarter Efficient Power Systems, First Edition. Edited by Zita Vale, Tiago Pinto, Michael Negnevitsky, and Ganesh Kumar Venayagamoorthy.

3.1 Introduction

The ever-increasing development of applications in the Internet space has led to an explosive growth of data scale in the digital world. This is occurring due to the fast development of technologies such as cloud computing, big data, and the Internet-of-Things (IoT) [1]. In comparison, global data in 2020 was expected to be 22 times that of 2011, reaching about 35.2 zettabytes, according to a report by the International Data Corporation in 2012 [2]. This boom of digital data brings huge opportunities and transformative potential for a wide variety of fields, such as in the healthcare industry, energy sector, enterprises, manufacturing, and educational services [3, 4]. As a result of the information overload across fields, artificial intelligence (AI) techniques such as deep learning (DL) have made breakthroughs in natural language processing, speech recognition, image processing, forecasting, and many other fields [5].

Technological companies such as Google, Apple, Twitter, and Facebook collect and analyze huge amounts of data, being one of the biggest investors in projects related to DL. For instance, Google employs DL algorithms on colossal amounts of data acquired from the Internet, for their services, such as Google Translator, Google Street View, the image search engine, and Android speech recognition [6]. Furthermore, Apple's virtual assistant, Siri, provides a wide range of services such as personalized news, reminders, and weather forecasts, among others, by combining DL and the increasing amount of available data [7].

DL approaches are also fundamental for dealing with rising relevant problems with high impact in societies, such as those related to power and energy systems (PES) [8]. Environmental problems and rising energy prices have contributed to a surge of research regarding energy consumption. Accordingly, renewable energy resources are fundamental for grid stability as well as their environmental benefits against climate change and global warming [9]. Renewable energy is expected to account for 40% by 2040 and accounts for 25% of today's electricity production [10]. As a result, there is an abundance of DL works that aim at forecasting photovoltaic (PV) energy, electricity prices, and wind power.

When applied to energy management systems (EMSs), DL can bring many benefits both in energy use and cost reductions, in residential spaces and workplaces. Several important works related to EMSs can be found in the literature, addressing topics such as heating, ventilation, and air conditioning (HVAC) systems, building energy consumption predictors, power quality disturbances detection in smart grids, fault detection (FD) and classification on generators, automatic feature engineering, energy management operational control, demand response (DR), transactions in an electricity market, among others [11].

In 1943 it was first published the McCulloch–Pitts model [12] by Warren McCulloch and Walter Pitts, the first artificial neuron, achieved by analyzing and summarizing the characteristics of neurons. Further research that led to neural networks was conducted in Ref. [13], which focused on a cell assembly theory in order to explain the adaptation of cerebral neurons during the learning process. Only in 1958 was the first perceptron invented by Frank Rosenblatt in the proposed work [14], which consisted of a kind of supervised learning with binary classifiers. In 1969, Marvin Minsky and Seymour Papert wrote an article citing that a perceptron could not learn the XOR function, which lead to a stagnation of neural network research for almost a decade. However, in 1982 it was proposed the Hopfield network [15], which led to a small resurgence in the field. The revival of artificial neural networks (ANNs) was done by David Rumelhart, Geoffrey Hinton, and Ronald Williams with the proposed paper [16], which introduced back-propagating in the learning process of neural networks, through a gradient descent method. Also noteworthy is the work done by David Ackley on Boltzmann machines by using simulated annealing with the proposed paper in

Ref. [17]. ANNs hit another nadir due to the appearance of various shallow machine learning (ML) approaches that had advantages both in theory and application. The concept of DL was only introduced after Geoffrey Hinton put forward the concept in 2006 in the journal *Science* [18], starting an ever-increasing trend of deep ANN research and application in the real world.

The first relevant works in DL consist of deep kernel machines for meta-heuristic searching [19], autoencoders (AEs) for acoustic processing [20], a recurrent neural network (RNN) using long short-term memory (LSTM) for natural language processing [21], a restricted Boltzmann machine (RBM), and a deep belief network (DBN) applied to robotics and linguistics [22], a DBN for image matching [23], a deep convolutional network for object recognition [24], a convolutional DBN for cluster formation and learning [25], and a stacked AEs for denoising [26].

In the field of PES, some of the first neural network approaches [27] propose an ANN to correct the results of a neuro-fuzzy forecast model to improve wind power forecasting, [28] employs an ANN for long-term electricity demand forecasting, [29] uses a dynamic RNN to optimize long-term energy performance forecasting of integrated generation systems, Mellit [30] proposes an RNN for daily PV generation forecasting, and Refs. [31, 32] applies an ANN model to predict electricity prices.

Considering the recent trend and relevance of DL models, this chapter provides an overview of the most relevant and recent research on DL in PES applications, addressing works related to regression, classification, and decision-making problems. The reviewed works consider models that provide contributions to the evolution of EMSs, energy resources forecasting, DR, electricity market, and power quality, among other relevant topics in the PES field.

For the present chapter literature review, several inclusion and exclusion criteria were considered, as shown in Tables 3.1 and 3.2, respectively. As a result, there are a total of 67 publications included in the review, from 2016 to 2021, as represented in Figure 3.1. Of all the publications included, 26 depict regression problems, 24 represent classification problems, and 17 describe decision-making problems, accounting for, 39%, 36%, and 25% of the overall publications included, respectively. Also noteworthy is that there is a larger number of publications included in the review that represent more recent studies (i.e. from 2019 to 2021), as shown in Figure 3.2. It demonstrates that in 2016, 2017, 2018, 2019, 2020, and 2021, there are a total of 2, 6, 10, 18, 16, and 15 publications included in the review, respectively. Therefore, even though the present review includes publications from 2016, most of the publications are from the most recent state-of-the-art literature.

This chapter structure is divided into four main sections. The first section introduces and contextualizes the objective of the chapter and the associated history regarding DL, and the specific application in PES. Section 3.2 presents the state-of-art literature review regarding DL in PES by subdividing it into regression, classification, and decision-making problems. Section 3.3 discusses the accomplishments, limitations, and challenges of DL applications in PES. Finally, Section 3.4 summarizes the chapter by describing the most promising and expected future advancements in the field.

Table 3.1 Inclusion criteria.

	Inclusion criteria
IC1	The source focuses on the application in intelligent power and/or energy systems
IC2	The source belongs to the field of deep learning in computer science
IC3	The source is peer-reviewed, except in cases that the source shows a relevant contribution to the area
IC4	The source describes a significant contribution to the fields of study

Table 3.2 Exclusion criteria.

	Exclusion criteria
EC1	The source is a duplicate
EC2	The source is published before 2016
EC3	The source is not written in English
EC4	The source is either a book chapter, dissertation, review, or thesis
EC5	The source does not present studies related to deep learning in intelligent power and/or energy systems

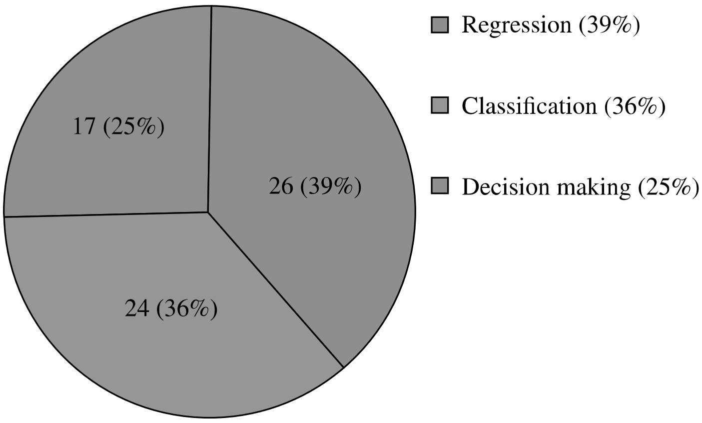

Figure 3.1 Number and percentage of publications per category (i.e. regression, classification, and decision-making) included in the review.

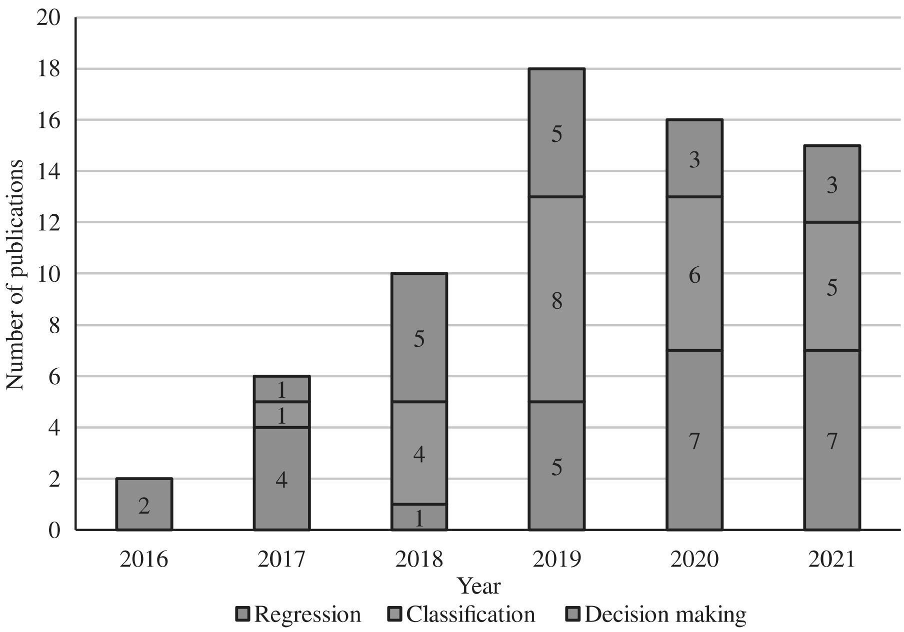

Figure 3.2 Number of publications per year included in the review.

3.2 Deep Learning

DL is a complex ML approach for pattern and structure recognition in very large datasets. These computational models are composed of multiple processing layers that allow visual object recognition, speech recognition, data forecast, and solutions in many other domains, to be greatly improved when compared to other ML approaches [5].

When considering PES, DL is widely used both on regression and classification problems, having a wide range of applications on smart grids, microgrids, smart buildings, energy market, and much more [11]. The regression problems addressed in this domain are mainly related to PV forecast [33], electricity prices forecast [34], predicting power flows [35], energy consumption prediction [36], abnormal energy consumption detection [37], and wind power forecast [38]. Regarding classification problems in PES, these are less common, but they still play a big role in detecting power quality disturbances [39], FD [40], and feature engineering in EMSs [41]. Furthermore, there are also decision-making problems that rely heavily on deep reinforcement learning (DRL), such as energy management [42], DR [43], and electricity market models [44], among others.

Sections 3.2.1–3.2.3 explore the existing solutions for these types of problems in PES, i.e. regression, classification, as well as decision-making problems. For each subchapter, the most relevant works in the literature are reviewed, with a focus on their contributions to PES.

3.2.1 Regression Problems

Sections 3.2.1.1–3.2.1.4 describe the most common regression problems in PES that are addressed using DL, by reviewing the most relevant and recent related works.

3.2.1.1 Photovoltaic Energy Forecast

The increasing penetration of renewable energy sources, and the associated uncertainty brought by the dependency on natural conditions such the wind and solar intensity, required the development of new advanced models that are able to provide good forecasts for this type of generation. Regarding PV generation, in specific, in Ref. [33] it is proposed a novel DL architecture, presented as PV-Net, which is capable of short-term forecasting of day-ahead PV energy. The PV-Net architecture focuses on the redesign of the gates on the gated recurrent unit (GRU) using convolutional layers, originating a so-called convolutional GRU. Furthermore, each convolutional GRU cell is stacked with each other in bidirectional blocks. This innovative approach allows for efficient extraction of temporal and positional characteristics in the PV power sequences, while also avoiding information loss across layers and improving gradient flow during training.

Also regarding PV forecast, Luo et al. [45] proposes an innovative DL approach that incorporates domain knowledge into LSTM to forecast the hourly day-ahead PV power generation. The proposed solution aims at overcoming the shortcoming of ML algorithms that are only dependent on large quantities of data. This solution proposes a physics-constrained long short-term memory, named PC-LSTM, composed of three different modules that represent the imposed constraints/knowledge: the data filtering module, the clipping module, and the loss penalty module. The first module focuses on the world/general knowledge of PV, by filtering unreasonable forecasts, for example, high PV generation during the night. The second module aims to represent natural science knowledge, by restricting the output to a reasonable range. For PV generation in specific, according to the physical law, the output needs, e.g. to be greater or equal to zero. Finally, the last module penalizes the LSTM whenever there are constraints that are being violated.

The work presented in Ref. [46] focuses on identifying the best algorithm for predicting solar energy. It compares three different approaches: an RNN, an LSTM, and a GRU. Moreover, the proposed case study uses real meteorological data, from 2016 to 2018. Results found that the RNN and LSTM have really good performance for this type of problem, showing very high levels of accuracy and low errors, with RNN coming slightly on top in terms of performance. On the other hand, the proposed GRU showed worse performance, lagging behind both RNN and LSTM.

Other noteworthy works on this topic are [47] which focuses on a hybrid DL approach, based on wavelet packet decomposition and LSTM, for short-term PV power forecasting, and Ref. [48] which combines LSTM, stacked AE, DBN, and multilayer perceptron techniques to forecast solar power in a smart grid.

PV forecasting approaches using DL have the ability to deal with the uncertainty related to this type of generation. Some of the considered advances are promising regarding the application to the forecasting of other types of variable generation. However, there are specific characteristics, e.g. concerning wind power forecast, which require the development of specific dedicated models [49].

3.2.1.2 Wind Power Forecast

The work of Wang et al. [38] proposes a novel DL-based ensemble approach for probabilistic wind power forecast. The proposed approach uses wavelet transform, CNN, and ensemble techniques to achieve a system capable of advanced point forecast. The wavelet transform method is used to decompose the raw wind power data into different frequencies; the nonlinear features are then used on the CNN for training. Furthermore, in order to deal with uncertainties in wind power data, i.e. data noise and model misspecification, these are identified separately in a later stage.

In Ref. [50], it is proposed a bidirectional LSTM model for wind power forecast of an Urban Regional Energy Internet. By using a bidirectional LSTM model, the proposed solution is capable of having better accuracy than a normal LSTM, since the latter only considers previous information, while the bidirectional LSTM can comprehensively consider historical and future information.

3.2.1.3 Building Energy Consumption Prediction

Not only renewable generation is associated with significant uncertainty, but also energy consumption [51]. It is therefore necessary that the prediction of energy consumption is enhanced by new models, benefiting from the advances in DL. For example, the work in Ref. [36] aims to create a model capable of predicting the energy consumption of a building while taking into account the randomness and noisy disturbances in data in this type of problem. The proposed approach combines extreme learning machine and stacked AEs, to create a novel approach called extreme stacked AE. The extreme learning machine is used to predict the building's energy consumption, while the stacked AEs are utilized to extract the building energy consumption features. Moreover, the usage of multiple AEs as building blocks, in the stacked AE, allows for the construction of a deep neural network (DNN). From the results of the proposed model, the paper claims that the proposed extreme stacked AE, when compared to other recent techniques, performs best when dealing with uncertainty in building energy consumption data.

Another building energy consumption predictor is proposed in Ref. [52], which focuses on comparing different models for predicting consumers' energy consumption in a grid. The validated methods are an ANN, support vector machine (SVM), RNN, and two newly developed stochastic models, specifically a conditional RBM and a factored conditional RBM. Results from the study showed that the factored conditional RBM outperforms every other method tested.

An innovative DL model for predicting energy consumption is proposed in Ref. [53] which incorporates IoT networks. The proposed solution called Energy-Net is composed of multiple stacked spatial-temporal modules, each one consisting of a temporal transformer and spatial transformer submodules. The temporal transformer submodule's purpose is to learn and extract long-range and short-range sequential energy characteristics along with the temporal domain. The spatial transformer submodule focuses on extracting hidden spatial information by integrating convolutional layers. Moreover, the later submodule includes an improved self-attention mechanism that has been demonstrated to be effective in learning the interactions among different elements.

Also on this topic, Berriel et al. [54] proposes a monthly energy consumption predictor that utilizes an LSTM with data normalization, using a standardization technique, and by utilizing metadata. When compared to other baseline methods that are employed by power companies, and models such as fully connected neural network, and convolution neural network (CNN), the proposed model by the paper outperforms all of them, confirming that LSTMs are more suitable for temporal task analysis.

Building energy consumption forecasting can also benefit from disaggregating consumption data to enable predictions of consumption per device, especially considering devices with large associated consumption and variation, such as HVAC and heat pumps. A DL approach using a feed-forward neural network with back propagation to predict energy consumption of HVAC systems is proposed in Ref. [55]. The study proposes a new variable selection process that aims at improving the model's accuracy when compared to other more traditional ML and statistical approaches. It was found that operational variables in the heating and cooling system, for example, the temperature at the top of the compressors, high pressure, low pressure, etc., were more important than temperature variables, such as outdoor temperature and indoor temperature.

An abnormal energy consumption detection system for ground source heat pumps is proposed in Ref. [37]. The proposed solution works by utilizing a mode decomposition-based LSTM to predict energy consumption, and then through the difference between the predicted and actual values, the anomaly is detected, by using the statistical model Grubbs' test. The system is capable of detecting three categories of anomalies: parabola anomaly, abrupt anomaly, and time-related anomaly.

3.2.1.4 Electricity Price Forecast

On the market side, Marszałek and Burczynski [34] proposes a novel approach to forecasting hourly day-ahead electricity prices, through the combination of multiple techniques, such as RNN, LSTM, attention mechanism, and clustering. Contrary to other proposed approaches in recent years, the paper claims that it focuses on the construction of a fair and unbiased model, which aims to increase accuracy and decrease categorical bias between the different clusters. To achieve such claims, the proposed model calculates the attention weights of the LSTM hidden states by considering a context vector given for each sample individually as the cluster center to which the sample belongs.

An innovative methodology for day-ahead electricity price forecast is presented in Ref. [56]. The proposed method utilizes Bayesian DL techniques for probabilistic energy price forecast and, in order to guarantee scalability in complicated network architectures, a special training method that focuses on processing mini-batches in parallel. Furthermore, the proposed model supports heteroscedasticity, eliminating the conventional homoscedastic assumption and the associated preprocessing cost. To validate the solution, case studies were conducted on two separate day-ahead marketplaces with varied behavior. Results show that the proposed methodology is capable of achieving strong performance in out-of-sample settings while also offering forecast uncertainty signals.

Another day-ahead electricity price forecasting system is proposed in Ref. [57] based on a novel hybrid DL framework. It considers the residuals between the actual electricity price and forecasting electricity price to improve training performance, uncertainties in electricity price, and it proposes a new feature preprocessing approach that allows for better identification of correlated features and outliners. The proposed methodology is composed of four modules: feature preprocessing module, DL-based point prediction module, error compensation module, and probabilistic prediction module. The first module is used to discover outliers and select the associated features of the electricity price series, through the usage of the isolation forest algorithm and the least absolute shrinkage and selection operator method. Then, in the point prediction module, complex nonlinear features are extracted, achieved by combining RNN, LSTM, DBN, and CNN. In the error compensation module, the residual error between the predicted and actual electricity prices is further reduced if possible. Finally, in the probabilistic prediction module, a quantile regression approach is used to quantify uncertainty at varying degrees of confidence. From the results, it is clear that the proposed methodology has better performance in point and probabilistic forecasting, when compared to methods such as light gradient boosting machine, back-propagation neural networks, support vector regression, and k-nearest neighbor.

3.2.1.5 Other Regression Works

Sections 3.2.1.1–3.2.1.4 discuss the most relevant works related to the most widely addressed regression problems in PES. Nevertheless, there are other, less explored regression problems, which are also noteworthy. The work in Ref. [58] proposes a distributed deep network structure for power system security knowledge discovery based on multitask learning to monitor and control power grids. The solution proposed aims to keep power systems secure, stable, and economical. A DBN is implemented to nonlinearly extract deep and abstract features layer-by-layer for total transfer capability regression tasks. Furthermore, it is proposed a new distributed training algorithm for the deep structure to accelerate the training process.

An optimal load dispatch with DL in a community microgrid is proposed in Ref. [59], by taking into account solar power and load demand forecast. In this study, a novel deep RNN with a LSTM model is implemented in a community microgrid to forecast aggregated power load and PV energy over a short-term period. According to the present study, the proposed deep RNN with a LSTM model has fewer limitations in modeling complex nonlinear problems and time dependencies. Therefore, having the potential for making more accurate short-term forecasting to support economical load dispatch systems, when compared to more traditional forecasting methods. Moreover, an optimal load dispatch model for grid-connected community microgrid considers residential power load, PV energy, electric vehicles, and energy storage systems. For load dispatch optimization, a particle swarm optimization algorithm is proposed.

For predicting optimal power flows on air conditioners, Fioretto et al. [35] proposes an innovative DL approach to the problem, by combining a feed-forward DNN and the Lagrangian dual method. The DNN is used to predict the generator setpoints for the optimal power flow of the air conditioner, while the Lagrangian dual method exploits the physical and engineering constraints by using violation degrees.

Other noteworthy works in the field, relating to regression problems in DL, are presented in Table 3.3.

3.2.2 Classification Problems

Sections 3.2.2.1–3.2.2.3 discuss the most prevalent classification problems in PES using DL, by reviewing the most recent and relevant research.

Table 3.3 Summary of other existing work for regression problems concerning DL approaches.

Article	Objective[a)]	Method[b)]	Application[c)]
Soleimanzade and Sadrzadeh [60]	ES, PVF	LSTM, CNN	Smart grid
Pramono et al. [61]	LDF, DR	LSTM, CNN	Smart grid
Wu and Wang [62]	AN, CN	FNN	Microgrids
Huang et al. [63]	EO	CDNN	MIMO-NOMA
Bugbee et al. [64]	UI, IV, S	DNN	ESM

a) ES: energy storage, PVF: photovoltaic forecast, LDF: load demand forecast, AN: action network, CN: critic network, EO: energy optimization, UI: user interface, IV: immersive visualization, S: statistics.
b) CDNN: communication deep neural network, FNN: feed-forward neural network.
c) MIMO-NOMA: multiple-input-multiple-output non-orthogonal multiple access, ESM: energy system models.

3.2.2.1 Power Quality Disturbances Detection/Classification

The paper in Ref. [39] implements a novel stacked AE for feature reduction and power quality disturbances detection and classification in the industrial sector. The proposed solution is able to classify multiple combinations of disturbances at the same time, up to 17 were validated in the paper. It works by feature extracting using a domain fusion approach, in order to characterize the electrical power grid. Then, using a stacked AE, an adaptive pattern characterization is carried out. Finally, a neural network structure is used to identify disturbances.

A sequence-to-sequence DL model is proposed in Ref. [65] that aims at detecting the type and time location of power quality disturbances. It achieves these results by using a bidirectional GRU. The input sequence is first normalized and batched. Then, deep features are retrieved from the input sequence using a bidirectional GRU RNN with several layers stacked together in both the forward and backward directions. Afterward, based on the retrieved features, a fully connected neural network layer and Softmax are used to determine the associated probability indicating the category to which each element in the input sequence is categorized. The Argmax or Top K operation is used to identify the category of each element in the input sequence by picking the highest likelihood. Finally, the category is recognized, and the starting and ending times of disturbances are also located just as the category is changed. Contrary to existing sequence-to-sequence models that employ an encoder-decoder network, the solution proposed in the paper is capable of capturing intrinsic temporal information of power quality disturbances, mainly the starting and ending times of disturbances.

Also regarding power disturbances detection, Mohan et al. [66] proposes an innovative CNN-LSTM to characterize and classify power quality disturbances. The proposed hybrid model contains two CNN layers followed by max pooling in order to extract low-level spatial dependencies among the data. Then, an LSTM layer extracts the long-term temporal dependencies and generates a more abstract feature map. Finally, a fully connected layer with a softmax function is used to calculate the probability of each class of disturbance. Furthermore, the proposed solution was compared to other known models, such as CNN, RNN, identity-RNN, LSTM, and GRU, demonstrating that it outperforms all of them. Also, it was found to be accurate for real-time characterization and classification of power quality disturbances.

In Ref. [67] it is proposed a new classification approach for power quality complex disturbances by combining an independent component analysis and a sparse AE. Also, the solution allows the detection and classification of multiple disturbances at the same time, with high accuracy. The sparse AE is used to extract the value of the features of a single disturbance. In addition, the proposed sparse AE is designed for multiple-label power quality disturbances classification, mainly for double and quadruple disturbances.

A new approach using DBN is proposed in Ref. [68] to classify power quality disturbances. The DBN proposed in the paper consists of several stacked RBMs for unsupervised learning that focuses on classifying seven types of disturbance signals: interruption, sag, swell, harmonic, oscillatory, sag-harmonic, and swell-harmonic. Results show that the proposed approach achieves much better classification rates than more conventional algorithms, such as wavelet transform and SVM.

Other relevant works that address problems related to power quality are Sindi et al. [69] which proposes the combination of 1D and 2D CNN features for power quality analysis, Ref. [70] aims to detect frequency disturbances through two CNNs and classifier fusion, and in Ref. [71] it is proposed an automatic system for complex power quality disturbance identification based on a multi-fusion CNN.

Power quality is hugely influenced by faults in the system. Fault detection and classification are, thereby, a significant problem that required dedicated solutions. For instance, condition monitoring, through FD and classification, in smart grids can be a great way to improve the overall power quality in the grid and reduce power quality disturbances [72].

3.2.2.2 Fault Detection/Classification

A DL-based fault detection and classification system for real-time diagnosis of air-handling units are proposed in Ref. [40]. It focuses on improving the operational efficiency of air-handling units to reduce the energy consumption of HVAC systems in buildings. In addition, the solution was validated using the EnergyPlus simulation platform in order to reduce the complexity of data processing and effectively ensure data integrity, thus improving the reliability of the proposed diagnostic model. With the usage of a DNN with back propagation, the proposed solution is capable of detecting up to five different fault types, operating in real-time, and providing better diagnostic ability than other more conventional approaches.

The work in Ref. [73] proposes a fault diagnosis system for wind turbine gearboxes, by using a stacked AE and SVM for frequency analysis. The stacked AE-based multiclass SVM classifier is used to learn higher level representation of input features in order to classify the fault. The proposed solution in the paper not only outperforms traditional multiclass SVM methods but is also nonintrusive to wind turbines. Also noteworthy, the solution has no or very low hardware cost since it does not require additional sensors that other traditional methodologies might need.

An intelligent fault diagnosis system using DL for solid oxide fuel cell systems is proposed in Ref. [74], which aims to successfully detect multiple faults simultaneously. To achieve these results, a novel simultaneous fault diagnosis model based on a stacked sparse AE is implemented. The proposed model automatically captures essential features from the original input data, thus reducing time on heavily hand-crafted features. Also, the proposed methodology does not require extensive knowledge of the physical model, thus having great generalization capabilities. When compared to other classical data-based diagnosis methods, the proposed solution has better performance on solid oxide fuel cell systems. In fact, the process of capturing essential features from the original input data as well as creating new features through the combination of original features is a relevant topic that enables the improvement of the learning process. Feature engineering works are providing significant contributions in this domain [75].

Other works worth noting regarding this topic are [76] which focuses on FD, classification, and location prediction in power systems by using an LSTM model, Afrasiabi et al. [77] proposes an FD system that integrates a CNN and a light-GRU applied to power transformers, and finally, a power cables FD and classification, by using a DBN, is proposed in Ref. [78].

3.2.2.3 Feature Engineering

In Ref. [41] it is proposed an innovative automatic feature engineering method using DL models to derive high-quality features for building energy prediction. It aims at reducing data dimensionality, decreasing prediction model complexity, and tackling the problem of corrupted and noisy data; thus improving building energy prediction performance. Three types of models are proposed to achieve such results, a fully connected AE, a convolutional AE, and a generative adversarial network. In addition, the models are compared to conventional feature engineering methods, for example, principal components and summarizing statistics. From the results, DL methods outperform all the conventional feature engineering methods, with generative adversarial networks coming on top of the DL models. In short, the generative adversarial network model, proposed in the paper, for both the generator and discriminator uses a convolutional AE as the basic operation unit. Also, since the discriminator only aims to distinguish whether a sample is real or synthetic, the output layer has only one neuron and the *sigmoid* function for classification.

The model proposed by Wang and Chen [79] focuses on a novel full closed-loop approach to detect and classify power quality disturbances through automatic feature extraction. To achieve these results, a CNN is designed to capture multi-scale features and reduce overfitting. Furthermore, the proposed closed-loop feedback, contrary to conventional methods which are divided into three phases: signal analysis, feature selection, and classification, allows for a unified whole that can be solved automatically by a universal DNN. When compared with other algorithms such as wavelet transform, SVM, particle swarm optimization, etc., under no noise, low-level noise, and high-level noise conditions, the accuracy of the proposed method outperforms all other algorithms; thus having a better noise immunity compared to conventional methods.

Another relevant work on this topic is presented in Ref. [80] which focuses on a multi-energy load prediction model. The main contribution of this work is the usage of a hybrid model, based on CNN and GRU, for effective feature extraction and dynamic modeling of time series.

3.2.2.4 Other Classification Works

In the paper [81] an innovative approach is introduced with aims at integrating DL and IoT for energy consumption reduction, by controlling the air conditioners. A people detection system, using a camera, is proposed in order to count the number of persons in a specific area, through a known CNN-based object detection network called You Look Only Once, abbreviated as YOLO, version 3. Accordingly, based on the number of people in the area, the air conditioners are optimized cost- and energy wise. However, the proposed work is designed to recognize human faces, instead of other more privacy-oriented approaches, which is a big red flag since it is to be implemented in homes and workplaces.

A novel architecture for feature learning and classification of voltage dips is proposed in Ref. [82]. An LSTM model is employed to extract voltage dip features from root-mean-square voltage sequences. Then, through a fully connected neural network layer with a *softmax* activation function, voltage dips are classified according to seven different types of dips, by using the feature provided by the LSTM outputs. Therefore the proposed solution is able to learn three-phase voltage dip patterns from time-dependent root-mean-square sequences. Contrary to conventional ML methods, the solution proposed in the paper allows learning dip features without requiring human knowledge or expert rules, transition-event segmentation, and selecting thresholds, while having better performance.

Table 3.4 Summary of other existing work for classification problems concerning DL approaches.

Article	Objective[a)]	Method[b)]	Application[c)]
Li et al. [83]	ETD	CNN-RF	Smart grid
Lu et al. [84]	FD, ETD	SSAE	Smart grid
Feng et al. [85]	BOD	CNN, BiLSTM	SB

a) ETD: electricity theft detection, BOD: building occupancy detection.
b) CNN-RF: convolutional neural network random forest, SSAE: semi-supervised autoencoder, BiLSTM: bidirectional long short-term memory.
c) SB: smart buildings.

Other noteworthy works in the field, relating to classification problems in DL, are described in Table 3.4.

3.2.3 Decision-Making Problems

Although regression and classification problems are the most widely addressed types of problems in DL within PES, DL is also used to address decision-making problems in this domain. Sections 3.2.3.1–3.2.3.3 present the most relevant and recent state-of-art literature review of decision-making problems in PES using DL.

3.2.3.1 Energy Management

An EMS using reinforcement learning based on Deep Q Network (DQN) is proposed in Ref. [42] for cost minimization on hybrid systems, which provide electricity, heating, cooling, and water for homes. The proposed method does not require online trial and error or historical data of system performance to determine the operational strategy. Furthermore, the developed reinforcement learning DQN takes into account hybrid systems with PV collectors, energy storage collectors, wind turbines, combined heating and power units, water heaters, hot water storage tanks, and batteries. When compared to two rule-based methods, Load Following and Full State of Charge, the proposed novel model outperforms both of them.

The paper in Ref. [86] proposes a real-time EMS to reduce energy costs of energy purchased for a storage PV system in a microgrid. To achieve this, a novel approach to the problem is presented, in which the focus of the cost optimization is on the grid energy purchases rather than on a direct storage control. A DRL is employed with a Q-learning algorithm to learn the optimal decision policy. Furthermore, many enhancements to improving learning speed and stability are implemented, that rely on experience replay, target network, and increasing discount factor.

Reference [87] focuses on a real-time dynamic EMS that considers the uncertainties of renewable energy availability and load demand, in a wind-solar-diesel-battery-reverse osmosis hybrid energy system. The proposed solution utilizes DRL with proximal policy optimization to obtain the optimal control policy that aims at reducing the overall energy costs. The proposed DRL method is a centralized control approach, having a single-point and communication delay problem.

An innovative real-time autonomous energy management approach for residential multi-energy systems is proposed in Ref. [88]. It uses a DRL-based approach combined with a novel deep deterministic policy gradient (DDPG) method with a prioritized experience replay strategy. Through the proposed approach, the agent is able to receive accurate feedback concerning its impact on energy management decisions in the multi-energy systems, thus allowing for better discovery of more cost-effective control strategies by exploring the entire action domain. In addition, the state-of-the-art DDPG allows for performance improvements when compared to state-of-the-art DRL methods, traditional stochastic programming, and robust optimization.

As shown by these works, today's energy management models need to deal with multiple variables, including variable renewable generation, electricity market prices, and uncertain consumption. It becomes essential that such systems are able to incorporate and motivate the means to reach suitable flexibility from the consumer side. In this context, DR becomes a central topic of research.

3.2.3.2 Demand Response

A model-free DRL method for DR management of interruptible load is proposed in Ref. [43]. The proposed DRL approach is combined with a Dueling DQN structure, in order to optimize DR management of interruptible load under the time of use tariff and variable electricity consumption patterns. An automatic DR architecture provides the possibility of real-time application of DR. In addition, the problem at hand is addressed as a Markov decision process (MDP), that is, the state, action, and reward function are defined, to obtain the maximum long-term profit of the DR management through the analysis of cumulative reward. The proposed architecture, when compared to other traditional DQN, overcomes the noise and instability. Moreover, the proposed DRL realizes the goal of minimizing the peak load demand and operational costs on the premise of regulating voltage to the safe limit.

A novel multi-agent DRL for DR participation in discrete manufacturing systems is presented in Ref. [89], which aims at minimizing electricity costs and improving grid stability. To achieve this goal, the industrial manufacturing system is designed as a partially observable Markov game. Also, in order to obtain the optimal schedule, for energy consumption, of the different machines, a multi-agent DDPG algorithm is proposed.

The successful widespread application of DR is not only dependent on the quality of the models that are used to manage it, but also on the intrinsic flexibility of consumers and on the incentives that are provided so that consumers are motivated to act in a flexible way. Such incentives are often associated with electricity market prices, which makes it crucial that suitable models are proposed to predict market prices, propose novel efficient market mechanisms, and support players' actions in the market.

3.2.3.3 Electricity Market

In Ref. [44] an innovative DRL with a replay mechanism for an event-driven local energy market is proposed. It aims at facilitating energy trading between prosumers at the distribution level. It uses a deep Q-learning algorithm for local energy trading modified from a DQN, to facilitate decision-making in the energy system as well as promote participation among prosumers. Furthermore, the trading process of a prosumer is modeled as an MDP to consider the volatile market and physical conditions.

Also regarding electricity markets, Xu et al. [90] proposes a DRL approach using a DDPG to solve an MDP for the optimal bidding in a wholesale electricity market and the energy price charged in the retailed electricity market for a load-serving entity. The MDP is formulated for the joint bidding and pricing problem, with continuous state and action spaces where the energy price and bid are two actions that share a common goal. Furthermore, a DNN is applied in order to reduce the significant number of state transitions that the DDPG algorithm requires, thus reducing application costs. The DNN learns from historical data, in order to model the wholesale electricity market and end-use customers for dynamic bid and price responses, respectively. Results have shown that the

Table 3.5 Summary of other existing work for decision-making problems concerning DL approaches.

Article	Objective[a)]	Method[b)]	Application[c)]
[93]	EC	DRL, DQN	Micro grid
[94]	CS	DRL, Q-learning	Smart grid
[95]	ED	DRL, Q-learning	ESU, DGU
[96]	ED, KT	DRL, Q-learning	Smart grid
[97]	EO	DRL, DQN, GDPG	Smart grid
[98]	OC	DRL, Q-learning	Wind turbine
[99]	OC	DRL, A3C	HVAC

a) EC: edge computing, CS: cyber security, ED: economic dispatch, KT: knowledge transfer, EO: energy optimization, OC: operational control.
b) GDPG: Gibbs deep policy gradient, A3C: asynchronous advantage actor critic.
c) ESU: energy storage units, DGU: distributed generation units.

proposed methodology makes more profit using the bidding and pricing policies, however, aggregate energy consumptions are reduced significantly.

3.2.3.4 Other Decision-Making Works

Further relevant decision-making related works can be found in the literature, for instance, the work, Diao et al. [91] proposes a novel framework, Grid Mind, for autonomous grid operational control using DRL. It proposes an intelligent agent that learns voltage control policies through interactions with offline simulations and is capable of adapting to new changes in load/generation variations and topological changes. To achieve this, a DQN is employed.

For optimal HVAC and window control strategy, Chen et al. [92] presents an approach using DRL to maximize natural ventilation. The proposed DRL utilizes a model-free Q-learning algorithm that focuses on minimizing both energy consumption and thermal discomfort by controlling HVAC and window systems. It takes into consideration temperature, humidity, solar radiation, and wind speed, both for indoor and outdoor environments. The proposed solution, when compared to a rule-based heuristic control strategy, outperforms it in all aspects, from lower energy consumption to lower discomfort degree hours and fewer high humidity hours.

Other noteworthy works in the field, relating to decision-making problems in DL, are represented in Table 3.5.

3.3 Accomplishments, Limitations, and Challenges

Although still a rather recent technology, DL is already bringing significant benefits to PES. DL forecasting accuracy has become one of the biggest accomplishments in the solution of PES-related problems. DL models provide forecasting capabilities that greatly outperform all simpler models and traditional ML techniques. LSTMs are at the forefront of this development, however many of them are hitting a wall in accuracy. Therefore, in recent years, new LSTMs combined with other methods and/or architectures are being proposed, such as the works described in Refs. [34, 38, 45, 54]. In addition, other approaches that aim at outperforming LSTMs were presented by Refs.

[33, 38, 53]. Furthermore, there is also a variety of new approaches being proposed for power quality disturbance and FD, that aim at detecting and classifying multiple combinations of disturbances or faults simultaneously, while retaining high levels of accuracy. These works are presented in Refs. [39, 66–68, 74]. Also noteworthy, are the novel automatic feature engineering methods in Refs. [41, 79]. Finally, the increasing usage of DRL for energy management has allowed real-time control systems to considerably minimize energy costs and consumption in buildings. Moreover, they take into consideration the usage of renewable energies to further reduce costs and outperform traditional optimization models with ease, such as in Refs. [42, 43, 86–88]. The good performance of DL models in the PES domain relies mainly on the ability to deal with large datasets, being able to extract relations between different data features and temporal patterns. Such a process is difficult, if not impossible when using simpler models. The complexity of DL approaches enables learning much more complex tasks and connections, being therefore a hugely promising approach for the future, as sustained by the already achieved results.

On the backside of the reviewed works, some papers do not focus on comparing their work to other state-of-the-art DL approaches; instead, they compare the proposed models to more conventional ML algorithms or traditional optimization algorithms. The problem with this approach is that DL relies heavily on performance to showcase its relevance, while aspects such as the potential application in other fields, the explainability, applicability to small datasets, etc., should be considered secondary. Therefore, it is unfair to compare two algorithms using different architectures based only on performance. These papers should focus more on comparing, for example, a DBN with a CNN for classification performance, and not with an SVM, as done in Refs. [55, 59, 68], as it is known that application requirements, e.g. size of the data set, are very different. Another shortcoming is the lack of real-world deployment or case studies, mainly in DRL methods. Works such as Refs. [43, 44, 86, 87, 89, 90, 92], etc. only showcase their results in a simulated environment, while no research is done for when the proposed solution is deployed in a real-world environment, where uncertainty and errors in data are common. Such assessment could improve the paper's reputation by comparing the results from the simulated and real-world environments. Moreover, in PES there is a big gap in the number of papers related to regression and classification problems. The majority of DL models in PES literature are focused on forecasting and consumption prediction methods, while a minority of research is done on the classification of power quality disturbances, FD, or other classification problems. This difference is felt when comparing the proposed works. For example, regression papers normally present new architectures or combinations of DNNs such as in Refs. [33–35, 38, 45, 53]; while classification works typically only explore different applications using already existing DL architectures, for instance in Refs. [40, 73, 81, 82], not innovating in DL techniques.

The extraction of explanations from DL algorithms is one of the biggest challenges in the upcoming years. The main problem of DL is that, while it shines by having the best performance among other ML techniques, DNNs represent "*black boxes*" and cannot explain why they came up with a number, classification, or decision [100]. Achieving explainable AI in DNNs allows for EMSs to provide the reasons for their actions. It would allow answering questions such as why was the air conditioner turned off, why does the forecast indicate that there will be an increase in electricity prices, or why is this power quality disturbance classified as flicker. As a result, trust in these systems would increase among consumers, leading to a possible rise in the usage of EMSs in residential spaces and workplaces. In addition, these systems could be used by field experts to verify if the DNNs are following the correct logic and associations. Another challenge is the incorporation of human expertise in DNNs. Integrating prior knowledge into DL systems is not as easy as it seems, primarily because DNNs represent knowledge through correlations between features, rather than abstractions such as in quantified statements [101]. Some works have started to tackle this problem, such as in Ref. [102], however with little progress. Achieving success in this area would allow for expert knowledge, for example, heuristic or physics laws, to be embedded into EMSs using DL. Subsequently, performance should increase as well as stability, as it would be able to deal more easily with problems outside of the training dataset. The integration of explainable AI with embedded human expertise would allow the management of knowledge in EMSs, providing experts and consumers better control of their "*black boxes*."

3.4 Conclusions

In this chapter, the current state-of-the-art in DL literature for intelligent PES has been reviewed. An introduction is made to contextualize the goals of the chapter and the evolution of DL since it first emerged. Afterward, the most recent and promising works in regression, classification, and decision-making problems using DL are presented and explored. These works demonstrate an ever-increasing trend of research in this area that aims at employing new models for particular domains, which outperform the more general conventional ML methods.

New approaches and methods for DL forecasting have become one of the biggest accomplishments, with many new methods outperforming traditional models and even the first architectures of LSTMs, according to various metrics. The detection and classification of multiple power quality disturbances and system faults, simultaneously, have also become a hot topic, as they retain levels of accuracy on par with single disturbance/fault classifiers. Novel automatic feature engineering methods are also being proposed, which aim at improving performance in PES-related problems. Furthermore, a wide variety of innovative DRL approaches are being proposed for energy management, taking into account renewable energy availability, energy storage, heating and power units, water heaters, and much more.

In the upcoming years, explainable AI and embedded human expertise in EMSs are going to be essential to increase trust among system users. However, there are still questions that need to be answered such as how far is humanity open to having fully AI-controlled homes and workplaces, as well as intelligent automatic energy trading and purchase between homes and industries.

References

1 Bhattarai, B.P., Paudyal, S., Luo, Y. et al. (2019). Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions. *IET Smart Grid* 2 (2): 141–154. https://doi.org/10.1049/IET-STG.2018.0261.

2 Gantz, B.J., Reinsel, D., and Shadows, B.D. (2012). Big data, bigger digital shadows, and biggest growth in the far east executive summary: a universe of opportunities and challenges. *IDC iView: IDC Analyze the Future* 2007: 1–16.

3 Gill, S.K., Nguyen, P., and Koren, G. (2009). Adherence and tolerability of iron-containing prenatal multivitamins in pregnant women with pre-existing gastrointestinal conditions. *Journal of Obstetrics and Gynaecology* 29 (7): 594–598. https://doi.org/10.1080/01443610903114527.

4 Manyica, J., Chui, M., Brown, B. et al. (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute (Issue May). https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation.

5 Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521 (7553): 436–444. https://doi.org/10.1038/nature14539.

6 Jones, N. (2014). Computer science: the learning machines. *Nature* 505 (7482): 146–148. https://doi.org/10.1038/505146a.

7 Efrati, A. (2013). *How "Deep Learning" Works at Apple, Beyond – The Information*. http://www.theinformation.com/articles/how-deep-learning-works-at-apple-beyond.

8 Hossain, M.S. and Mahmood, H. (2020). Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast. *IEEE Access* 8: 172524–172533. https://doi.org/10.1109/ACCESS.2020.3024901.

9 Abrishambaf, O., Lezama, F., Faria, P., and Vale, Z. (2019). Towards transactive energy systems: an analysis on current trends. *Energy Strategy Reviews* 26: 100418. https://doi.org/10.1016/j.esr.2019.100418.

10 Vural, G. (2020). Renewable and non-renewable energy-growth nexus: a panel data application for the selected Sub-Saharan African countries. *Resources Policy* 65: 101568. https://doi.org/10.1016/j.resourpol.2019.101568.

11 Runge, J. and Zmeureanu, R. (2021). A review of deep learning techniques for forecasting energy use in buildings. *Energies* 14 (3): 608. https://doi.org/10.3390/en14030608.

12 McCulloch, W.S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics* 5 (4): 115–133. https://doi.org/10.1007/BF02478259.

13 Shaw, G.L. (1986). Donald Hebb: the organization of behavior. In: *Brain Theory*, 231–233. https://doi.org/10.1007/978-3-642-70911-1_15.

14 Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review* 65 (6): 386–408. https://doi.org/10.1037/H0042519.

15 Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America* 79 (8): 2554–2558. https://doi.org/10.1073/PNAS.79.8.2554.

16 Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature* 323 (6088): 533–536. https://doi.org/10.1038/323533a0.

17 Ackley, D.H., Hinton, G.E., and Sejnowski, T.J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science* 9 (1): 147–169. https://doi.org/10.1016/S0364-0213(85)80012-4.

18 Hinton, G.E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation* 18 (7): 1527–1554. https://doi.org/10.1162/NECO.2006.18.7.1527.

19 Bengio, Y. and LeCun, Y. (2007). Scaling learning algorithms towards AI. In: *Large-Scale Kernel Machines* (ed. L. Bottou, O. Chapelle, D. DeCoste, and J. Weston), 182–192. MIT Press. https://nyuscholars.nyu.edu/en/publications/scaling-learning-algorithms-towards-ai.

20 Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P.A. (2008). Extracting and composing robust features with denoising autoencoders. In: *Proceedings of the 25th International Conference on Machine Learning*, 1096–1103. https://doi.org/10.1145/1390156.1390294.

21 Collobert, R. and Weston, J. (2008). A unified architecture for natural language processing: deep neural networks with multitask learning. In: *Proceedings of the 25th International Conference on Machine Learning*, 160–167. https://doi.org/10.1145/1390156.

22 Le Roux, N. and Bengio, Y. (2008). Representational power of restricted Boltzmann machines and deep belief networks. *Neural Computation* 20 (6): 1631–1649. https://doi.org/10.1162/NECO.2008.04-07-510.

23 Salakhutdinov, R. and Murray, I. (2008). On the quantitative analysis of deep belief networks. In: *ICML '08: The 25th Annual International Conference on Machine Learning held in conjunction with the 2007 International Conference on Inductive Logic Programming*, Helsinki Finland (5–9 July 2008). New York, NY, United States: Association for Computing Machinery. https://doi.org/10.1145/1390156.

24 Jarrett, K., Kavukcuoglu, K., Ranzato, M., and LeCun, Y. (2009). What is the best multi-stage architecture for object recognition? In: *Proceedings of the IEEE International Conference on Computer Vision*, 2146–2153. https://doi.org/10.1109/ICCV.2009.5459469.

25 Lee, H., Grosse, R., Ranganath, R., and Ng, A.Y. (2009). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: *ICML '09: The 26th Annual International Conference on Machine Learning held in conjunction with the 2007 International Conference on Inductive Logic Programming*, Montreal, Quebec, Canada (14–18 June 2009). New York, NY, United States: Association for Computing Machinery, pp. 609–616. https://doi.org/10.1145/1553374.1553453.

26 Vincent, P., Larochelle, H., Lajoie, I. et al. (2010). Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research* 11: 3371–3408.

27 Katsigiannis, Y.A., Tsikalakis, A.G., Georgilakis, P.S., and Hatziargyriou, N.D. (2006). Improved wind power forecasting using a combined neuro-fuzzy and artificial neural network model. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3955. LNAI, *SETN '06: Proceedings of the 4th Helenic conference on Advances in Artificial Intelligence*, Heraklion Greece (18–20 May 2006). Berlin, Heidelberg: Springer-Verlag, pp. 105–115. https://doi.org/10.1007/11752912_13.

28 Al Mamun, M. and Nagasaka, K. (2005). Artificial neural networks applied to long-term electricity demand forecasting. In: *HIS '04: Proceedings of the Fourth International Conference on Hybrid Intelligent Systems*, Kitakyushu, Japan (5–8 December 2004). NW Washington, DC, United States: IEEE Computer Society, pp. 204–209. https://doi.org/10.1109/ichis.2004.27.

29 Bonanno, F., Capizzi, G., and Tina, G. (2009). Long-term energy performance forecasting of integrated generation systems by recurrent neural networks. In: *ICCEP 2009 : International Conference on Clean Electrical Power*, Capri, Italy (9–11 June 2009). NW Washington, DC, United States: IEEE Computer Society, pp. 673–678. https://doi.org/10.1109/ICCEP.2009.5211956.

30 Mellit, A. and Shaari, S. (2009). Recurrent neural network-based forecasting of the daily electricity generation of a Photovoltaic power system. *Ecological Vehicle and Renewable Energy (EVER)*, Monaco (26–29 March 2009).

31 Jiang, H. and Tan, Z. (2012). Load forecasting in demand response. In: *2012 Asia-Pacific Power and Energy Engineering Conference (APPEEC 2012)*, Shanghai, China (27–29 March 2012). NW Washington, DC, United States: IEEE Computer Society. https://doi.org/10.1109/APPEEC.2012.6307716.

32 Pinto, T., Morais, H., and Corchado, J.M. (2019). Adaptive entropy-based learning with dynamic artificial neural network. *Neurocomputing* 338: 432–440. https://doi.org/10.1016/j.neucom.2018.09.092.

33 Abdel-Basset, M., Hawash, H., Chakrabortty, R.K., and Ryan, M. (2021). PV-Net: an innovative deep learning approach for efficient forecasting of short-term photovoltaic energy production. *Journal of Cleaner Production* 303: 127037. https://doi.org/10.1016/j.jclepro.2021.127037.

34 Marszałek, A. and Burczynski, T. (2021). Forecasting day-ahead spot electricity prices using deep neural networks with attention mechanism. *Journal of Smart Environments and Green Computing* 21–31. https://doi.org/10.20517/jsegc.2021.02.

35 Fioretto, F., Mak, T.W.K., and Van Hentenryck, P. (2020). Predicting AC optimal power flows: combining deep learning and Lagrangian dual methods. *Proceedings of the AAAI Conference on Artificial Intelligence* 34 (01): 630–637. https://doi.org/10.1609/aaai.v34i01.5403.

36 Li, C., Ding, Z., Zhao, D. et al. (2017). Building energy consumption prediction: an extreme deep learning approach. *Energies* 10 (10): 1525. https://doi.org/10.3390/en10101525.

37 Xu, C. and Chen, H. (2020). Abnormal energy consumption detection for GSHP system based on ensemble deep learning and statistical modeling method. *International Journal of Refrigeration* 114: 106–117. https://doi.org/10.1016/j.ijrefrig.2020.02.035.

38 Wang, H.Z., Li, G.Q., Wang, G.B. et al. (2017). Deep learning based ensemble approach for probabilistic wind power forecasting. *Applied Energy* 188: 56–70. https://doi.org/10.1016/j.apenergy.2016.11.111.

39 Gonzalez-Abreu, A.-D., Delgado-Prieto, M., Osornio-Rios, R.-A. et al. (2021). A novel deep learning-based diagnosis method applied to power quality disturbances. *Energies* 14 (10): 2839. https://doi.org/10.3390/en14102839.

40 Lee, K.P., Wu, B.H., and Peng, S.L. (2019). Deep-learning-based fault detection and diagnosis of air-handling units. *Building and Environment* 157: 24–33. https://doi.org/10.1016/j.buildenv.2019.04.029.

41 Fan, C., Sun, Y., Zhao, Y. et al. (2019). Deep learning-based feature engineering methods for improved building energy prediction. *Applied Energy* 240: 35–45. https://doi.org/10.1016/j.apenergy.2019.02.052.

42 Sanaye, S. and Sarrafi, A. (2021). A novel energy management method based on Deep Q Network algorithm for low operating cost of an integrated hybrid system. *Energy Reports* 7: 2647–2663. https://doi.org/10.1016/j.egyr.2021.04.055.

43 Wang, B., Li, Y., Ming, W., and Wang, S. (2020). Deep reinforcement learning method for demand response management of interruptible load. *IEEE Transactions on Smart Grid* 11 (4): 3146–3155. https://doi.org/10.1109/TSG.2020.2967430.

44 Chen, T. and Su, W. (2018). Local energy trading behavior modeling with deep reinforcement learning. *IEEE Access* 6: 62806–62814. https://doi.org/10.1109/ACCESS.2018.2876652.

45 Luo, X., Zhang, D., and Zhu, X. (2021). Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge. *Energy* 225: 120240. https://doi.org/10.1016/j.energy.2021.120240.

46 Jebli, I., Belouadha, F.Z., Kabbaj, M.I., and Tilioua, A. (2021). Deep learning based models for solar energy prediction. *Advances in Science, Technology and Engineering Systems* 6 (1): 349–355. https://doi.org/10.25046/aj060140.

47 Li, P., Zhou, K., Lu, X., and Yang, S. (2020). A hybrid deep learning model for short-term PV power forecasting. *Applied Energy* 259: 114216. https://doi.org/10.1016/j.apenergy.2019.114216.

48 Gensler, A., Henze, J., Sick, B., and Raabe, N. (2017). Deep learning for solar power forecasting – an approach using autoencoder and LSTM neural networks. In: *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Budapest, Hungary (9–12 October 2016). NW Washington, DC, United States: IEEE Computer Society, pp. 2858–2865. https://doi.org/10.1109/SMC.2016.7844673.

49 Sogabe, T., Ichikawa, H., Sakamoto, K. et al. (2016). Optimization of decentralized renewable energy system by weather forecasting and deep machine learning techniques. In: *IEEE PES*

Innovative Smart Grid Technologies Conference Europe, 1014–1018. https://doi.org/10.1109/ISGT-Asia.2016.7796524.

50 Weng, G., Pei, C., Ren, J. et al. (2021). Modeling and forecasting of wind power output of urban regional energy internet based on deep learning. *Journal of Physics: Conference Series* 1732 (1): 12190. https://doi.org/10.1088/1742-6596/1732/1/012190.

51 Nam, K.J., Hwangbo, S., and Yoo, C.K. (2020). A deep learning-based forecasting model for renewable energy scenarios to guide sustainable energy policy: a case study of Korea. *Renewable and Sustainable Energy Reviews* 122: 109725. https://doi.org/10.1016/j.rser.2020.109725.

52 Mocanu, E., Nguyen, P.H., Gibescu, M., and Kling, W.L. (2016). Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks* 6: 91–99. https://doi.org/10.1016/j.segan.2016.02.005.

53 Abdel-Basset, M., Hawash, H., Chakrabortty, R.K., and Ryan, M. (2021). Energy-net: a deep learning approach for smart energy management in IoT-based smart cities. *IEEE Internet of Things Journal* https://doi.org/10.1109/JIOT.2021.3063677.

54 Berriel, R.F., Lopes, A.T., Rodrigues, A. et al. (2017). Monthly energy consumption forecast: a deep learning approach. In: *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, Alaska (14–19 May 2017). NW Washington, DC, United States: IEEE Computer Society, pp. 4283–4290. https://doi.org/10.1109/IJCNN.2017.7966398.

55 Hwang, J.K., Yun, G.Y., Lee, S. et al. (2020). Using deep learning approaches with variable selection process to predict the energy performance of a heating and cooling system. *Renewable Energy* 149: 1227–1245. https://doi.org/10.1016/j.renene.2019.10.113.

56 Brusaferri, A., Matteucci, M., Portolani, P., and Vitali, A. (2019). Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. *Applied Energy* 250: 1158–1175. https://doi.org/10.1016/j.apenergy.2019.05.068.

57 Zhang, R., Li, G., and Ma, Z. (2020). A deep learning based hybrid framework for day-ahead electricity price forecasting. *IEEE Access* 8: 143423–143436. https://doi.org/10.1109/ACCESS.2020.3014241.

58 Huang, T.E., Guo, Q., Sun, H. et al. (2019). A deep learning approach for power system knowledge discovery based on multitask learning. *IET Generation, Transmission and Distribution* 13 (5): 733–740. https://doi.org/10.1049/iet-gtd.2018.5078.

59 Wen, L., Zhou, K., Yang, S., and Lu, X. (2019). Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting. *Energy* 171: 1053–1065. https://doi.org/10.1016/j.energy.2019.01.075.

60 Soleimanzade, M.A. and Sadrzadeh, M. (2021). Deep learning-based energy management of a hybrid photovoltaic-reverse osmosis-pressure retarded osmosis system. *Applied Energy* 293: 116959. https://doi.org/10.1016/j.apenergy.2021.116959.

61 Pramono, S.H., Rohmatillah, M., Maulana, E. et al. (2019). Deep learning-based short-term load forecasting for supporting demand response program in hybrid energy system. *Energies* 12 (17): 3359. https://doi.org/10.3390/en12173359.

62 Wu, N. and Wang, H. (2018). Deep learning adaptive dynamic programming for real time energy management and control strategy of micro-grid. *Journal of Cleaner Production* 204: 1169–1177. https://doi.org/10.1016/j.jclepro.2018.09.052.

63 Huang, H., Yang, Y., Ding, Z. et al. (2020). Deep learning-based sum data rate and energy efficiency optimization for MIMO-NOMA systems. *IEEE Transactions on Wireless Communications* 19 (8): 5373–5388. https://doi.org/10.1109/TWC.2020.2992786.

64 Bugbee, B., Bush, B.W., Gruchalla, K. et al. (2019). Enabling immersive engagement in energy system models with deep learning. *Statistical Analysis and Data Mining* 12 (4): 325–337. https://doi.org/10.1002/sam.11419.

65 Deng, Y., Wang, L., Jia, H. et al. (2019). A sequence-to-sequence deep learning architecture based on bidirectional GRU for type recognition and time location of combined power quality disturbance. *IEEE Transactions on Industrial Informatics* 15 (8): 4481–4493. https://doi.org/10.1109/tii.2019.2895054.

66 Mohan, N., Soman, K.P., and Vinayakumar, R. (2018). Deep power: deep learning architectures for power quality disturbances classification. In: *2017 International Conference on Technological Advancements in Power and Energy (TAP Energy)*, Kollam, Kerala, India (21–23 December 2017). NW Washington, DC, United States: IEEE Computer Society, pp. 1–6. https://doi.org/10.1109/TAPENERGY.2017.8397249.

67 Shi, X., Yang, H., Xu, Z. et al. (2019). An independent component analysis classification for complex power quality disturbances with sparse auto encoder features. *IEEE Access* 7: 20961–20966. https://doi.org/10.1109/ACCESS.2019.2898211.

68 Li, C.M., Li, Z.X., Jia, N. et al. (2018). Classification of power-quality disturbances using deep belief network. In: *2018 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, Chengdu, China (15–18 July 2018). NW Washington, DC, United States: IEEE Computer Society, pp. 231–237. https://doi.org/10.1109/ICWAPR.2018.8521311.

69 Sindi, H., Nour, M., Rawa, M. et al. (2021). A novel hybrid deep learning approach including combination of 1D power signals and 2D signal images for power quality disturbance classification. *Expert Systems with Applications* 174: 114785. https://doi.org/10.1016/J.ESWA.2021.114785.

70 Wang, W., Yin, H., Chen, C. et al. (2020). Frequency disturbance event detection based on synchrophasors and deep learning. *IEEE Transactions on Smart Grid* 11 (4): 3593–3605. https://doi.org/10.1109/TSG.2020.2971909.

71 Qiu, W., Tang, Q., Liu, J., and Yao, W. (2020). An automatic identification framework for complex power quality disturbances based on multifusion convolutional neural network. *IEEE Transactions on Industrial Informatics* 16 (5): 3233–3241. https://doi.org/10.1109/TII.2019.2920689.

72 Gonzalez-Abreu, A.D., Saucedo-Dorantes, J.J., Osornio-Rios, R.A. et al. (2020). Deep learning based condition monitoring approach applied to power quality. In: *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Vienna, Austria (8–11 September 2020). NW Washington, DC, United States: IEEE Computer Society, pp. 1427–1430. https://doi.org/10.1109/ETFA46521.2020.9212076.

73 Cheng, F., Wang, J., Qu, L., and Qiao, W. (2018). Rotor-current-based fault diagnosis for DFIG wind turbine drivetrain gearboxes using frequency analysis and a deep classifier. *IEEE Transactions on Industry Applications* 54 (2): 1062–1071. https://doi.org/10.1109/TIA.2017.2773426.

74 Zhang, Z., Li, S., Xiao, Y., and Yang, Y. (2019). Intelligent simultaneous fault diagnosis for solid oxide fuel cell system based on deep learning. *Applied Energy* 233, 234: 930–942. https://doi.org/10.1016/j.apenergy.2018.10.113.

75 Mohammadi, H.G., Arshad, R., Rautmare, S. et al. (2020). Deepwind: an accurate wind turbine condition monitoring framework via deep learning on embedded platforms. In: *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Vienna, Austria (8–11 September 2020). NW Washington, DC, United States: IEEE Computer Society, pp. 1431–1434. https://doi.org/10.1109/ETFA46521.2020.9211880.

76 Belagoune, S., Bali, N., Bakdi, A. et al. (2021). Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems. *Measurement* 177: 109330. https://doi.org/10.1016/J.MEASUREMENT.2021.109330.

77 Afrasiabi, S., Afrasiabi, M., Parang, B., and Mohammadi, M. (2020). Designing a composite deep learning based differential protection scheme of power transformers. *Applied Soft Computing* 87: 105975. https://doi.org/10.1016/J.ASOC.2019.105975.

78 Zhang, Y., Mei, W., Dong, G. et al. (2018). A cable fault recognition method based on a deep belief network. *Computers and Electrical Engineering* 71: 452–464. https://doi.org/10.1016/J.COMPELECENG.2018.07.043.

79 Wang, S. and Chen, H. (2019). A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network. *Applied Energy* 235: 1126–1140. https://doi.org/10.1016/j.apenergy.2018.09.160.

80 Xuan, W., Shouxiang, W., Qianyu, Z. et al. (2021). A multi-energy load prediction model based on deep multi-task learning and ensemble approach for regional integrated energy systems. *International Journal of Electrical Power & Energy Systems* 126: 106583. https://doi.org/10.1016/j.ijepes.2020.106583.

81 Elsisi, M., Tran, M.Q., Mahmoud, K. et al. (2021). Deep learning-based industry 4.0 and internet of things towards effective energy management for smart buildings. *Sensors (Switzerland)* 21 (4): 1–19. https://doi.org/10.3390/s21041038.

82 Balouji, E., Gu, I.Y.H., Bollen, M.H.J. et al. (2018). A LSTM-based deep learning method with application to voltage dip classification. In: *2018 18th International Conference on Harmonics and Quality of Power (ICHQP)*, Ljubljana, Slovenia (13–16 May 2018). NW Washington, DC, United States: IEEE Computer Society, pp. 1–5. https://doi.org/10.1109/ICHQP.2018.8378893.

83 Li, S., Han, Y., Yao, X. et al. (2019). Electricity theft detection in power grids with deep learning and random forests. *Journal of Electrical and Computer Engineering* 2019: https://doi.org/10.1155/2019/4136874.

84 Lu, X., Zhou, Y., Wang, Z. et al. (2019). Knowledge embedded semi-supervised deep learning for detecting non-technical losses in the smart grid. *Energies* 12 (18): 3452. https://doi.org/10.3390/en12183452.

85 Feng, C., Mehmani, A., and Zhang, J. (2020). Deep learning-based real-time building occupancy detection using AMI data. *IEEE Transactions on Smart Grid* 11 (5): 4490–4501. https://doi.org/10.1109/TSG.2020.2982351.

86 Kolodziejczyk, W., Zoltowska, I., and Cichosz, P. (2021). Real-time energy purchase optimization for a storage-integrated photovoltaic system by deep reinforcement learning. *Control Engineering Practice* 106: 104598. https://doi.org/10.1016/j.conengprac.2020.104598.

87 Zhang, G., Hu, W., Cao, D. et al. (2021). Data-driven optimal energy management for a wind-solar-diesel-battery-reverse osmosis hybrid energy system using a deep reinforcement learning approach. *Energy Conversion and Management* 227: 113608. https://doi.org/10.1016/j.enconman.2020.113608.

88 Ye, Y., Ye, Y., Qiu, D. et al. (2020). Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning. *IEEE Transactions on Smart Grid* 11 (4): 3068–3082. https://doi.org/10.1109/TSG.2020.2976771.

89 Lu, R., Li, Y.C., Li, Y. et al. (2020). Multi-agent deep reinforcement learning based demand response for discrete manufacturing systems energy management. *Applied Energy* 276: 115473. https://doi.org/10.1016/j.apenergy.2020.115473.

90 Xu, H., Sun, H., Nikovski, D. et al. (2019). Deep reinforcement learning for joint bidding and pricing of load serving entity. *IEEE Transactions on Smart Grid* 10 (6): 6366–6375. https://doi.org/10.1109/TSG.2019.2903756.

91 Diao, R., Wang, Z., Shi, D. et al. (2019). Autonomous voltage control for grid operation using deep reinforcement learning. In: *2019 IEEE Power & Energy Society General Meeting (PESGM)*, Atlanta, Georgia, United States of America (4–8 August 2019). NW Washington, DC, United States: IEEE Computer Society. https://doi.org/10.1109/PESGM40551.2019.8973924.

92 Chen, Y., Norford, L.K., Samuelson, H.W., and Malkawi, A. (2018). Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. *Energy and Buildings* 169: 195–205. https://doi.org/10.1016/J.ENBUILD.2018.03.051.

93 Munir, M.S., Abedin, S.F., Tran, N.H., and Hong, C.S. (2019). When edge computing meets microgrid: a deep reinforcement learning approach. *IEEE Internet of Things Journal* 6 (5): 7360–7374. https://doi.org/10.1109/JIOT.2019.2899673.

94 Ni, Z. and Paul, S. (2019). A multistage game in smart grid security: a reinforcement learning solution. *IEEE Transactions on Neural Networks and Learning Systems* 30 (9): 2684–2695. https://doi.org/10.1109/TNNLS.2018.2885530.

95 Liu, W., Zhuang, P., Liang, H. et al. (2018). Distributed economic dispatch in microgrids based on cooperative reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems* 29 (6): 2192–2203. https://doi.org/10.1109/TNNLS.2018.2801880.

96 Han, C., Yang, B., Bao, T. et al. (2017). Bacteria foraging reinforcement learning for risk-based economic dispatch via knowledge transfer. *Energies* 10 (5): 638. https://doi.org/10.3390/EN10050638.

97 Sogabe, T., Malla, D.B., Takayama, S. et al. (2018). Smart grid optimization by deep reinforcement learning over discrete and continuous action space. In: *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC)*, Waikoloa, Hawaii, United States of America (10–15 June 2018). NW Washington, DC, United States: IEEE Computer Society, pp. 3794–3796. https://doi.org/10.1109/PVSC.2018.8547862.

98 Saenz-Aguirre, A., Zulueta, E., Fernandez-Gamiz, U. et al. (2019). Artificial neural network based reinforcement learning for wind turbine yaw control. *Energies* 12 (3): 436. https://doi.org/10.3390/EN12030436.

99 Zhang, Z., Chong, A., Pan, Y. et al. (2018). A deep reinforcement learning approach to using whole building energy model for HVAC optimal control. In: *2018 Building Performance Analysis Conference and SimBuild* co-organized by ASHRAE and IBPSA-USA, Chicago, Illinois, United States of America (28 September 2018). NW Washington, DC, United States: IEEE Computer Society, pp. 675–682. https://www.researchgate.net/publication/326711617.

100 Holzinger, A. (2018). From machine learning to explainable AI. In: *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, Košice, Slovakia (23–25 August 2018). NW Washington, DC, United States: IEEE Computer Society, pp. 55–66. https://doi.org/10.1109/DISA.2018.8490530.

101 Marcus, G. (2018). Deep learning: a critical appraisal. CoRR. https://arxiv.org/abs/1801.00631v1.

102 Mitsuhara, M., Fukui, H., Sakashita, Y. et al. (2019). Embedding human knowledge into deep neural network via attention map. In: *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Volume 5 VISAPP: VISAPP, Online Streaming, 2021 (8–10 February 2021), pp. 626–636. https://arxiv.org/abs/1905.03540v4.