

Review

Review of Big Data Analytics for Smart Electrical Energy Systems

Huilian Liao ^{1,*}, Elizabeth Michalenko ¹ and Sarat Chandra Vegunta ²

¹ Power, Electrical and Control Engineering, Sheffield Hallam University, Sheffield S1 1WB, UK; e.michalenko@shu.ac.uk

² Siemens PTI, Siemens Industry, Inc., 400 State St., 4th Floor, Schenectady, NY 12305, USA; chandra.vegunta@siemens.com

* Correspondence: huilian.liao@shu.ac.uk

Abstract: Energy systems around the world are going through tremendous transformations, mainly driven by carbon footprint reductions and related policy imperatives and low-carbon technological development. These transformations pose unprecedented technical challenges to the energy sector, but they also bring opportunities for energy systems to develop, adapt, and evolve. With rising complexity and increased digitalization, there has been significant growth in the amount of data in the power/energy sector (data ranging from power grid to household levels). Utilization of this large data (or “big data”), along with the use of proper data analytics, will allow for useful insights to be drawn that will help energy systems to deliver an increased amount of technical, operational, economic, and environmental benefits. This paper reviews various categories of data available in the current and future energy systems and the potential benefits of utilizing those data categories in energy system planning and operation. This paper also discusses the Big Data Analytics (BDA) that can be used to process/analyze the data and extract useful information that can be integrated and used in energy systems. More specifically, this paper discusses typical applications of BDA in energy systems, including how BDA can be used to resolve the critical issues faced by the current and future energy network operations and how BDA contributes to the development of smarter and more flexible energy systems. Combining data characterization and analysis methods, this review paper presents BDA as a powerful tool for making electrical energy systems smarter, more responsive, and more resilient to changes in operations.



Citation: Liao, H.; Michalenko, E.; Vegunta, S.C. Review of Big Data Analytics for Smart Electrical Energy Systems. *Energies* **2023**, *16*, 3581. <https://doi.org/10.3390/en16083581>

Academic Editor: Silvio Simani

Received: 17 March 2023

Revised: 17 April 2023

Accepted: 19 April 2023

Published: 20 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: big data analytics; network planning and operation; smart electrical energy systems; demand side management; artificial neural networks; low-carbon technologies

1. Introduction

With increased digitalization in power grids, a large amount of data (or “big data”) will be generated and stored for current or future use. As the number of data sources increases, effective acquisition and processing of the data—gathered in large quantities via those sources individually and/or collectively—is of great importance to ensure reliability of the insights drawn from that data [1]. Characterized by the five V’s of big data [2], Big Data Analytics (BDA) in power systems draws data from various data sources and types to enable faster analyses of large quantities of that data. Data considered for BDA purposes can include typical power system measurements [3] as well as nontraditional data drawn from sources not intended for use in the power sector [4].

With the availability of big data in power grids, the usefulness of that data (typically raw data) to draw useful insights via BDA—with powerful big data processing capabilities—has attracted great interest among both academia and industry. Given the large amount of raw data to handle, the choice of applied BDA method for that data is important, as the benefits realized will significantly depend on the selected BDA method. Various

BDA methods have been developed, including artificial neural networks, deep learning, reinforcement learning, etc. [5,6].

As the operation and management of electrical energy systems grow more complex, it becomes more financially and functionally beneficial to utilize the new pool of ever-growing data sources and analysis methods. With applications across industries, BDA is presented in this paper as a powerful tool for power system operation and management. BDA's capabilities in working with vast datasets make its use in smart electrical energy systems an important industry development that will aid in the current and future transformations facing the grid. As such, discussions on BDA for smart electrical energy systems are timely and relevant to the evolution of smart energy systems.

With powerful capabilities of BDA in processing data, BDA presents great potential to assist in the planning and operation aspects of energy systems, particularly in the context of the great challenge current and future energy systems face in their transition to a net-zero future. Much of the Low Carbon Technology (LCT)-based equipment will be connected to, distributed across, and operating in modern power grids. Among LCTs, a significant number of electricity-based loads, including Electric Vehicles (EV) and electrical heat pumps, will be integrated into power systems in the very near future [7]. At the same time, conventional power generators, with large system inertia, are shutting down and are being replaced with small-system-inertia, LCT, and intermittent-energy-based generation systems. Driven by carbon footprint reduction goals and related policies, various sectors, including heavy industries, are increasingly becoming electrified, leading to an increased load burden and reliance on the current and future power systems for energy supply. Furthermore, due to increased electrification, including for transport via EVs, and increased generation from intermittent, LCT-based energy resources, large variations in system power flows, including conditions leading to thermal constraints, could be expected—such operations could exacerbate the already aging power system infrastructure, leading to increased system failures and faults. On the other hand, these LCTs are typically interfaced with power systems via power electronic systems, which have high controllability. Thus, these network changes allow for a high degree of flexibility to support effective network management and operation. In current power systems, flexibility resources (such as controllable loads, renewable energy generation, batteries, etc.) have been developed and are purchased to tackle issues that would otherwise require expensive, traditional network reinforcement methods. However, as flexibility resources are increasing in number, existing systems are not able to manage this large amount of flexibility resources and related data efficiently or optimally. In this regard, BDA plays an important role in supporting flexible resource management and smart system operation. BDA has been used to forecast load consumption and renewable energy generation under different network operating scenarios [8,9]. With an accurate forecast of system operation conditions, system operators can better understand their networks' operating conditions and develop effective strategies to optimally manage the available network resources or optimally purchase the required flexibility resources to achieve greater techno-economic benefits. BDA has been applied to resolve other critical issues, such as estimation and fault diagnosis, to ensure the power systems' stability [10].

This paper presents an overview of the data available from power systems and the potential analysis of that data, via BDA, to draw useful insights. In terms of big data characteristics and collection methods, previous reviews have differed in categorization and application. Detailed reviews of big data concepts have addressed data categorization and analysis methods across various applicable industries. Those regarding power systems have discussed an overview of data collection and its uses. This study details the BDA process through the expanded definition of data categorization for BDA purposes, the preprocessing methods that bridge data collection with analysis, and the complex technical analysis methods to provide a more detailed review of BDA's value specifically for power systems. Different from other review papers, this paper not only presents the state-of-the-art in BDA that can be applied in the electrical energy sector but also provides key

techniques that can be used to integrate BDA for power system applications. The techniques that are especially designed to tailor BDA for power contexts are critical for a successful application. The paper discusses the capabilities/benefits of using BDA to address critical issues, such as load forecasting, fault diagnosis, sag estimation, etc., faced by the current and future power systems as well as the potential barriers to full implementation of BDA and its benefits. This review paper consists of three sections that address data, BDA, and its applications, respectively. Section 2 introduces the data that are available in electrical energy systems. Section 3 introduces the BDA techniques that have great potential in power system applications; Section 4 presents several typical BDA applications that bring both data and BDA in the power system context, presenting the potential benefits it can bring to electrical energy systems.

2. Data in Electrical Energy Systems

Data are essential for modern power systems to function properly and for system planners and operators to complete critical functions. Modern data collection in a power system primarily comprises data from customer meters, Phasor Measurement Units (PMUs), utility Supervisory Control and Data Acquisition (SCADA) systems, etc. This data can be utilized to gather the real-time state of the grid as well as allow for formulating effective responses to managing system events or faults. As the number of data sources grows, effective acquisition and processing of large quantities of data becomes imperative for a reliable gathering of insights and for decision making [1].

When applying big data processes, nearly all data are considered useful, since complex analyses can quickly derive valuable insights from large amounts of varied inputs and data types. Therefore, data utilized for big data analysis range extensively, from real-time power system measurements to social media and user traffic as inputs [4]. Although there are no clear boundaries on what data can be utilized for big data in power systems, specific examples of data sources are further discussed in the later sections of this paper.

2.1. *V's of Big Data*

Big data are not inherently defined as large quantities of data but rather the distinct processes of extracting meaningful information from those large quantities [11]. In formal definitions, information associated with big data carries with it three key characteristics, the three V's of big data: volume, velocity, and variety [2]. The addition of the characteristics of veracity and value is a much-discussed modification to the formal definition [12]. As data collection equipment carries an inherent assumption of non-zero error that can affect function, both veracity and value are considered here as important factors for managing big data in a power system context.

2.1.1. Volume

The first V, volume, comprises the amount of data that is collected and analyzed. In typical big data applications, information is defined in the order of terabytes and petabytes, constituting massive quantities of data points that require advanced processing and storage capabilities [4]. Smart meters installed in homes are just one example of the data sources and the scale of data collection that such devices entail, generating about 120 GB of data points per day per meter [11]. The UK grid alone has almost 18 million smart or advanced electricity meters and has seen an average increase of over 2 million meters per year over the last five years [13]. Although much of the data available from meters and sensors is deleted locally if no significant event is detected [4], the potential applications of this information in BDA warrant greater storage capabilities.

While not all collected data can be stored and utilized due to the sheer volume and space constraints, the ability to store greater amounts of data enables greater pattern recognition and analytics. Therefore, the more data, the better. Research into data storage and processing capabilities seeks to address this disparity through some methods of paring down storage requirements, either by distributed data storage systems that use digital

networks to spread the load [3] or by processing collected information using data compression [14]. Regardless of the storage methods, the sheer quantity of data available from power system measurements, consumption patterns, and even weather forecasts contribute greatly to the applications of results drawn from BDA.

2.1.2. Velocity

The second V, velocity, comprises the speed of data generation and transmission of that data. The frequency of data collection directly affects the volume of data that is collected and that needs to be stored. Metering units and sensors have varying sampling rates, which range from readings at near-constant frequency to event-triggered readings, although much of the data are deleted locally if not immediately utilized [4]. Utility devices such as PMUs have sampling rates as high as 240 samples per second, while smart meters and satellite weather data can take as long as 15 min between measurements [11].

Since the conclusions reached by BDA are more effective with a greater quantity of data points, ideal data sources would sample more frequently, thus reducing the need for interpolation and estimation. However, while continuous data streams leave fewer gaps between points, the challenge of processing the data fast enough to warrant its collection begins to arise [11]. Various data optimization, compression, and searching methods for more advanced data analyses are crucial for enabling higher collection rates [12]. Once the data is collected, parsing through it quickly to classify and identify meaningful data among the collected data falls into the realm of advanced data mining processes [15].

2.1.3. Variety

The third V, variety, refers to the range of data that are collected and utilized. Historical power system data exist in a structured format and come directly from metering units and measurements [3]. However, BDA in a power system context relies heavily on the variety of data sources and types of data. Utility devices like smart meters and digital fault recorders source data alongside consumption trends and even satellite images [11]. As different measuring units and sources record data in various formats, big data processes can make use of collected information regardless of data structure [16]. Collected data are either structured, semistructured, or unstructured [17]. These data structures are explained further in later sections of this paper.

As collection systems and sources continue to evolve, the amount of semistructured and unstructured data is growing. Most notably, the use of information from sources not specific to or intended for use in the power sector is driving the rise in unstructured data [4]. While greater variety of data facilitates greater meaning, processing this semistructured and unstructured data is challenging for systems and methods historically reliant on the rigidity of structured data [11]. All information collected under various formats must be sorted and converted to usable structures for processing [16]. Preprocessing stages and methods are detailed in later sections of this paper.

2.1.4. Veracity

Although not included in the widely considered formal definition of big data, veracity is included as the fourth V in a power system context. Veracity refers to the accuracy and reliability of the data. As data collected from various sources carry with it the potential for error, veracity in data processing is imperative for extracting meaningful information. Large collections of data have great potential for inaccuracies due to measuring errors, equipment tolerances, or transmission noise [3]. For example, smart meters, relied upon for providing a detailed look at local metering points, have as much as a 2.5% measurement error [11]. While this alone is not a particularly high error, a system reliant on the readings of millions of smart meters, each with a 2.5% error, may not be able to optimally process and then respond to network states. Updates to both software and hardware to ensure accuracy will be essential as data sources become more varied [16].

2.1.5. Value

As with veracity, the fifth V, value, is added for consideration in a power system context, although its inclusion is not exclusive to BDA for smart energy systems. Value emphasizes the importance of extracting meaningful information from the collected data and is made possible by a large amount of data (volume) from a range of sources (variety) that is analyzed quickly (velocity) and accurately (veracity) [2]. As such, the value of the data is derived from the other four Vs's and is specific to the applications of what power system operators seek to optimize: generation mixes, load forecasting, economic constraints, energy consumption, etc. [11].

2.2. Data Types

Information collected and processed under the umbrella of BDA is categorized in the literature into two prevailing types: domain and off-domain data [4]. These two fields demonstrate a broad range of potential data sources that BDA can handle, reiterating the argument that there is no limit on useful data for BDA.

2.2.1. Domain Data

Domain data are defined as information specific to the power sector. In many cases, this data come directly from utility measurements and power systems' hardware: SCADA data as a near-continuous stream of information on the local grid states; consumption and usage patterns from smart meters; and graphical information from synchro-phasors. The information from utility measurements can be further broken down into operational and nonoperational data, classifying the purpose and end user of the collected data points [18]. Equally important domain data are that pertaining to economic dispatch, electricity markets, and pricing datasets [4]. Collectively, domain data constitute the whole of the power sector, with data sources ranging from generation to day-ahead market pricing.

In real-world applications, domain data are already in use for operational purposes and grid maintenance. SCADA data have been utilized in planning strategies for wind turbine maintenance and fault diagnosis [19]. Historically, domain data have been the standard source of meaningful information about the state of a power system.

2.2.2. Off-Domain Data

Off-domain data are defined as the information that is not specific to the power sector. Often, this data were not intended for use in the power industry, and they come from a wide range of sources nontraditionally utilized for energy management purposes. This can include information from social media, including videos and images, weather data, geographic information, traffic, and water usage [4]. Off-domain data can also be drawn from economic indicators outside of electricity markets.

The use of off-domain data in BDA for power systems is relatively new. Case studies have sought to demonstrate how these nontraditional data sources can be meaningful in curating valuable insights, including from the use of social media and online engagement, for smart energy systems and optimization of those systems [20,21]. Much of what constitutes off-domain data are considered mined data that arise from alternative networks and sources [4].

2.3. Data Structures

Due to a variety of data sources and formats, data systems must cope with data of varying types: structured, semistructured, and unstructured. While a majority of data are structured [4], the proliferation of nontraditional sensors and measurements drives an increase in alternative data structures, creating challenges for systems that reliably process structured data.

2.3.1. Structured

Structured data make up most data in power systems. Although structured data can come from either domain or off-domain sources, they are characteristically tabular- or database-driven. That is, structured data take the form of traditionally defined numeric datasets and have long been the cornerstone of data analytics [17]. In smart energy systems, structured data may be collected in the form of a table of continuously measured SCADA values and can include data on equipment, such as parameters, load control, distribution management, etc. [3].

2.3.2. Semistructured

Semistructured data are rising in prevalence in data analytics. Often in the form of XML or a similar format, these data lie at the edge between human and computer readability. They can also include information from legacy sources and equipment [17]. Therefore, semistructured data can take the form of load monitoring, power quality measurements, web data, geographical information, etc. [3].

2.3.3. Unstructured

Unstructured data greatly add to the complexity of data analytics and can be in the form of information with text, audio, or video [4]. Text mining is often used to convert the information into structured data that are more easily analyzed and processed [17]. In a power system context, this data may be collected from meteorological information, customer service interactions, etc. [3].

2.4. Data Preprocessing

As discussed in previous sections, the sheer volume of data sources and, consequently, data points presents great opportunities for BDA applications and the meaning derived from those analyses. Data collected from these various sources in raw form can be utilized for analysis. However, raw data are riddled with inconsistencies (redundancies, outliers, missing data points, etc.) that obscure meaningful data and decrease the accuracy of forecasting. To ensure maximum derived value from BDA methods, data preprocessing must be carried out [22]. Through preprocessing methods, datasets are clarified and reduced for more effective analysis.

Data analysis typically undergoes two major stages: acquisition and integration [22]. Additional preprocessing steps carried out before BDA analysis methods can aid in achieving more consistent and usable data and, therefore, greater derived meaning. This section combines typical data analysis stages and additional preprocessing steps to detail how raw data are modified for analysis.

2.4.1. Data Acquisition

Data acquisition refers to the gathering of data from their various available sources [22]. These sources are discussed in greater detail in the above sections. Acquired data are imperfect and carry with them an inherent nonzero error in the form of missing points, incorrect values, repeated measurements, outliers, etc. Data clean-up is carried out through data reduction and data cleansing.

Data reduction transforms vast datasets into a form that retains accurate values in a simplified structure [23]. Work in [24] demonstrates a filtering approach for selecting data points with relevant features and [25] describes an outlier rejection algorithm for identifying outlier values that may otherwise skew analysis. The removal of extreme or repeated values results in a reduced dataset that may be more easily and effectively analyzed via BDA methods.

Data cleansing identifies incorrect values in a dataset [26]. Unlike data reduction, data cleansing involves modifying existing data that may be inaccurate or corrupt. Data cleansing can be carried out by inserting data points where they are missing due to measurement or storage errors, but it can also be done to replace outlier values with more

stable values [22]. Filling in missing or outlier values can be completed using regression or imputation methods [22]. Additionally, a variety of filtering methods are utilized to cleanse signals and datasets of noise that may be present. By replacing inaccurate or problematic values, datasets become cleaner and more consistent for analysis purposes.

2.4.2. Data Integration

Data integration refers to the combining of data from multiple sources [22]. The variety of available sources is discussed in detail in the above sections. Data from various physical sources are stored in files and databases, which must then be integrated into single sources. While integration is necessary to combine disparate datasets, doing so increases the complexity, thus requiring the use of data reduction techniques [22].

Central to data integration is the process of data transformation in which data structures are modified to achieve greater comparability and, therefore, integration. Data transformation ensures that datasets are usable in future analyses without modifying the nature of the original data [22]. Normalization is one method of data transformation that commonly utilizes standard deviation and mean values to reliably standardize datasets with varying scales. The result of data integration and transformation is a consistent single source of values that can be easily analyzed via BDA methods.

3. Big Data Analytics in a Power System Context

BDA is a class of techniques used for processing data and performing functions such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Deep Neural Networks (DNN), Long Short-Term Memory (LSTM), etc. BDA has been widely applied in solving various complex problems, including problems such as classification, clustering, pattern recognition, predictive analysis, data forecasting, statistical analysis, natural language processing, etc. This section will introduce several key BDA techniques that have great application potential in a power system context.

3.1. Big Data Analytics

3.1.1. Artificial Neural Network (ANN)

ANN is among the most popular Artificial Intelligence (AI) approaches that was inspired by biological nervous systems, consisting of neurons and a web of their interconnections. Simple processing elements, emulating a typical biological neuron, and a network of those interconnected elements allow for identification of complex patterns within a set of data, provided as an input, and for storage and quick application of those identified patterns for future use. ANN is usually made of three layers, namely the input layer, output layer, and hidden layer. The input layer takes the data from the network and is connected to the hidden layer and then the output layer. Based on the interconnection of nodes and data movement between the layers, ANN can be classified into two basic categories: the feedforward ANN and the recurrent ANN. In the feedforward ANN, the information moves from input layer to output layer in one direction. The recurrent ANN allows some of the information to move in the opposite direction as well. Figure 1 presents a simplified structure of an ANN that consists of hidden and output layers. Each layer has a set of weights (W) to be assigned and to be applied to the input vector (p); then, a bias vector (b) is added upon the weighted input vector before it is fed into a transfer function (f) for further processing. The output of the transfer function, f , in the hidden layer serves as an input to the output layer; the results generated from the transfer function, f , in the output layer will be the final outputs of an ANN network. The ANN parameters, including weights and biases, are determined by training the ANN, involving a set of data (including both inputs and the expected outputs) fed as an input to the ANN and adjusting the ANN parameters until the ANN outputs match the outputs of the training dataset to a certain desired accuracy [5].

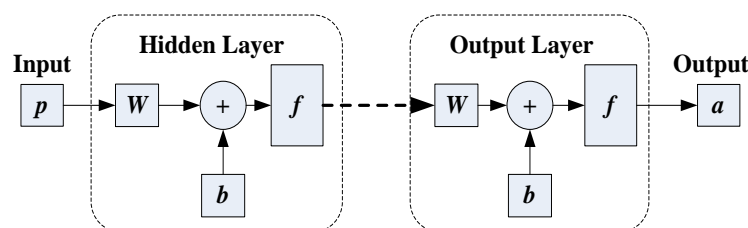


Figure 1. Illustration of the structure of an ANN.

ANN has been widely used for various applications in power systems that require identification of a hidden relationship between the known input–output datasets and then applying that identified relationship to an input dataset—with unknown outputs—to quickly predict outputs. With its merit of powerful data analytics, it has been also investigated for the applications of load forecasting, economic dispatch, fault diagnosis, harmonic analyzing, and system security assessment [27].

The ANN structure, such as the network size (e.g., the sizes for the input/output neurons) or the choice of transfer function, is not universal for all applications. The ANN structure should be selected and optimized for the needs of the specific application. With its merits of prediction accuracy and high adaptation, beyond applications in power systems, ANN has also attracted great attention in a wide range of applications in the areas of prediction, clustering, curve fitting, etc.

3.1.2. Deep Learning Techniques

Deep learning techniques have attracted a great deal of attention in the past decade [28] due to their promising results and great accuracy in large-scale pattern recognition, especially in solving visual recognition problems [6]. Deep learning techniques include Multilayer Perceptrons (MLPs), autoencoders, Convolutional Neural Networks (CNNs), LSTM, Recurrent Neural Networks (RNNs), etc.

MLP, as one of the early developed deep learning techniques, is a feedforward neural network with multiple perceptron layers equipped with activation functions. Both input layers and output layers in MLPs are Fully-Connected (FC). MLPs have been popularly applied to image and speech recognition problems.

Autoencoders mainly consist of three components: encoder, coder, and decoder. They are a specific type of feedforward neural network that have identical inputs and outputs. They have been mainly used for image processing, popularity prediction, etc.

CNNs consist of convolution layers that have a set of learnable filters/kernels that slide through input data to extract patterns in the data. Following the convolution layers, an FC neural network is used for classification. In both convolutional layers and neural network layers, multiple layers are typically used to model the complexity of patterns in large data. As one of the most popular deep learning techniques, CNN is popularly used for image recognition. RNNs, which have connections that form directed cycles, allow the outputs to be fed into hidden state as inputs, allowing for the previous inputs to be memorized. RNNs are commonly used for time-series analysis and for natural-language processing problems.

It has been found that deep learning techniques typically perform better than general neural networks, especially in solving multiple data class problems and machine learning applications with complex data structures. To cope with a large amount of data and to facilitate the extraction of more informative features [29], deep learning techniques usually require a relatively large number of hierarchical layers, suggesting the need for a large computational effort [30]. Deep learning techniques, due to the development in high-performance hardware, sometimes specifically designed for faster execution of those techniques, are becoming increasingly popular in various practical applications.

With its powerful pattern recognition capabilities and great accuracy in exacting information from complex data structure (formed by measurements collected from various locations and devices with different sampling frequency), deep learning has great potential in solving various power system problems. For instance, the unsupervised deep learning approach auto-encoder has been applied for load profile classification [31]. CNNs have been used to estimate the state-action value function for controlling residential load control [32], although its application at a system level is limited. Further exploration is needed to fully exploit the powerful capabilities of CNNs in pattern recognition and estimation applications.

3.2. Approaches to Integrate BDA in a Power System Context

Due to the large scale of actual power system/networks, it is impractical, if not impossible, or cost ineffective to have measurements at all desired locations. On the other hand, not all measurement data are useful to achieve a desired application's objective. Variables/parameters in BDA, therefore, should be carefully selected and processed to ensure usefulness of the collected data specific to a selected or prevailing system scenario.

Furthermore, power system performance highly depends on the topologies and operating conditions that vary constantly. It is desirable that the topology/configuration of power systems can be embedded in the input matrices of the BDA learning mechanisms. Per [33], the System Area Mapping process of feature extraction from the input data matrix was analyzed from a power system configuration perspective. The input matrix was arranged in a way that a patch in the input matrix summarizes the topology information of the corresponding area in the considered power system. Taking a 24-bus test network (given in Figure 2) as an example, different square patches in the input matrix map corresponding areas in the considered power system. Later, by sliding the kernels through these square patches, the features/characteristics in the local area can be extracted and integrated into a higher level in the feature map. Usually, many different kernels will be used in order to extract information from different aspects, capturing a varied set of characteristics and patterns that exist in a considered power system. Through this large set of kernels and through a number of feature extraction layers, useful information captured at a local area will eventually be summarized and integrated into a global area. These approaches should be tailored from a power system context, particularly where the performance of a considered technique as part of the BDA is greatly dependent on the considered power system, its configuration/topology, and its operating conditions.

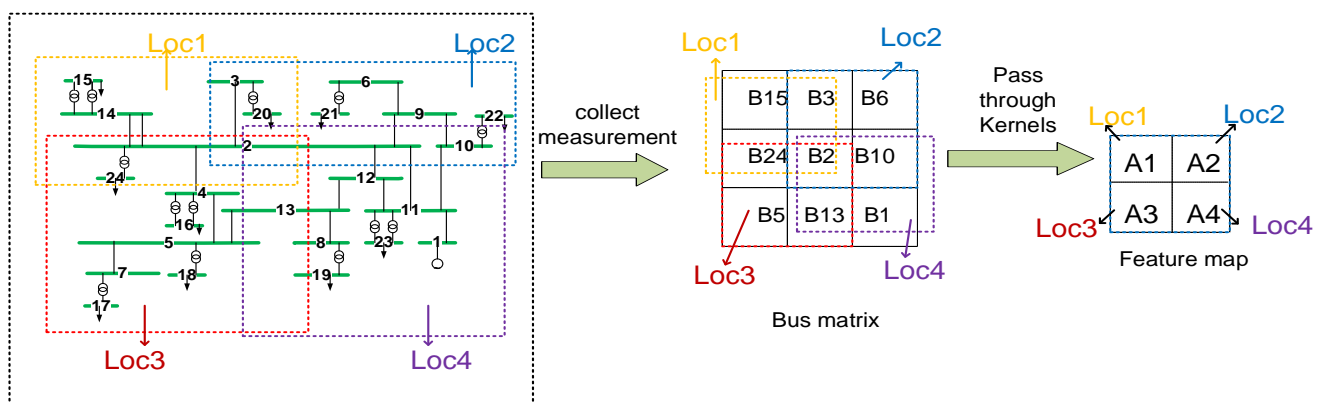


Figure 2. Illustration of system area mapping.

4. BDA Applications in Smart Power/Energy Systems

Traditional power system analysis and the related methodology development were mainly based on model-based analysis. With the fast Information Technology (IT) development and the development of powerful BDA and related tools, the integration of data-driven approaches started to attract great attention in both academia and industry.

A multidisciplinary approach of a hybrid of data-driven and model-based approaches is acknowledged to be the most effective way to resolve many power/energy issues [4].

Different power system applications have their unique characteristics. Depending on the nature of the applications, BDA cannot be transferred directly from one application to another, and it needs to be tailored and redesigned to adapt to the different applications that have different data sources, data structures, and objectives to achieve. Especially for the hybrid data-driven and model-based approach, the corresponding BDA solution/methods require a higher level of customization to adapt to applications.

A wide range of network issues can be resolved using BDA. In [34], large-scale Electric Vehicle (EV) charging management and operation in a grid environment, vehicle-to-grid (V2G), has been implemented using various BDA techniques, including DNN, LSTM, etc. Here, in [34], several typical power network problems are listed that can be successfully resolved to acceptable tolerances in industry.

4.1. Load Forecasting

Power system operation requires accurate prediction of future load demands (e.g., hourly, daily, etc. projections) to develop a better operating strategy to manage the system at that predicted time. With an accurate load forecast, the system operating condition will be known in advance, and a corresponding developed strategy can be used to optimally utilize the available resources in the system at the future time. In this way, the operation cost can be greatly reduced. In some cases, power system operators require a long-term load forecast that can assist with the long-term system planning process and allow for the determination of an optimal action plan and the required investment to accommodate the predicted load and/or generation growth at different locations in the system. Additionally, both active and reactive powers can be predicted to gather a more thorough understanding of the future network operating conditions, although the current interest has been focused more on the real power prediction alone.

Apart from forecasting the electrical power, in some applications, the expected load characteristics are also required. Based on different load characteristics, those loads can be classified into different categories (such as controllable or uncontrollable loads, from a load-controllability perspective). Furthermore, it will be beneficial for operators to receive information, such as the percentage of different load types (i.e., disaggregation of loads). This information will enable a better understanding of the load profiles and can be utilized to develop a better operating strategy. This detailed information on loads can be used in Demand Side Management (DSM) applications, which are increasingly used to support power systems' peak shaving applications. Figure 3 shows an example of the use of load forecasting in a DSM application. With a strategic operational plan, it helps system operators to put mechanisms in place for a stable power system during a DSM operation.

To resolve these forecast issues, various BDA methods have been proposed as follows: ANN [8,9], multiple regression [35], etc. The feasibility of ANN-based approaches for load forecasting has been validated via a few industrial applications. For example, in [36,37], an ANN-based approach has been used for reactive power prediction. The BDA-based method developed in [8,9,38] has been widely used for short-term load forecasting, with the mean forecast errors around 2% to 5%. The ANN-based approach proposed in [39] was able to achieve mean absolute percentage errors between 1% and 4%. The wavelet-based neural network developed in [40] has been used to forecast commercial loads (mainly focusing on short-term load forecasting), with the mean absolute percentage errors between 0% and 5%. The neural network proposed in [41] has been used to predict the load one or a few days ahead, with the mean error ranging between 1.5% and 3%. Based on the accuracy requirement, appropriate BDA methods can be selected and tailored for the specific application.

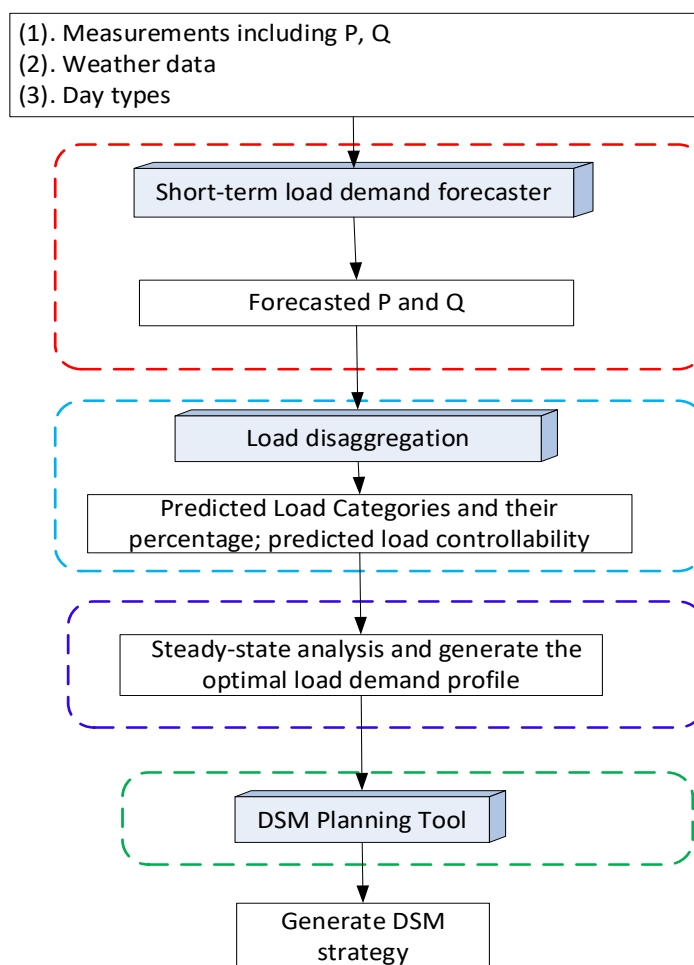


Figure 3. Illustration of load forecasting and DSM.

4.2. Fault or Outage Detection and Diagnosis

Fault diagnosis—due to the complexities of faults and their attributes (such as fault types, fault location, fault impedance, etc.) and together with the uncertainties of the resulting phenomena due to varying operating conditions—is among the most challenging issues in power systems.

ANN can be used to associate the network faults with the measured voltages at monitored buses. In this way, the faults can be identified once the measurements are received. In [10], ANN was applied to register faults, following the procedure detailed in Figure 4. As part of fault identification, the fault indices are, typically, arbitrarily given and they generally do not have a logical connection to faults. A choice of fault indexing system can affect the outputs of fault diagnostics. Thus, an optimal index allocation should be adopted to ensure accurate fault diagnostics and related outputs. In ANN applications, the neural network is trained to adapt to the training data given and fault indices (part of the training data) are fixed and not updated during training. In [10], an innovative approach has been proposed to choose the optimal strategy of fault indexing during training. In other words, both the neural networks and training data (fault indices) are updated simultaneously during the training procedure. It can be seen in Step 5, Figure 4 that the fault indices are renumbered based on the ranking of outputs in an ascending order. This way, the fault indices are renumbered by considering the magnitude of the simulated output of the trained neural network. This ensures the use of optimal fault indices and improves the final detection results. Even when limited monitored data are given, the approach in [10] is able to identify the faults accurately.

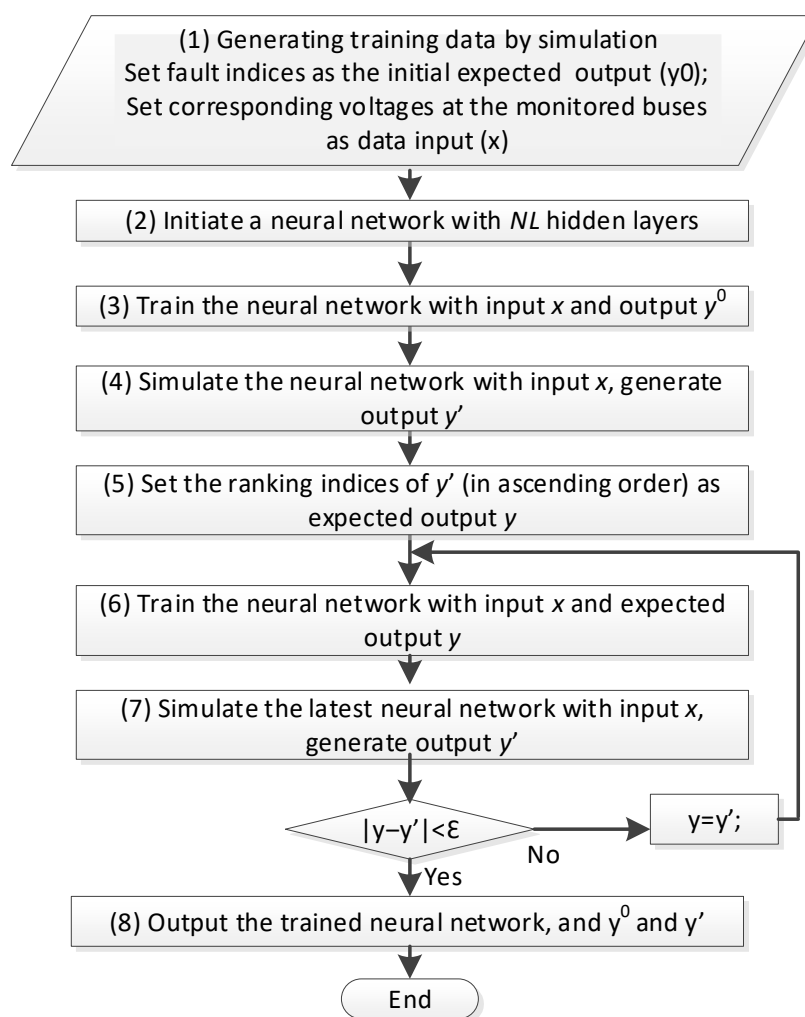


Figure 4. Procedure of obtaining the neural network and corresponding representative fault indices.

4.3. Voltage Sag Estimation

Voltage sags—due to their impact on both customers and utilities leading to significant financial losses—have attracted increased attention in both academia and industry. Voltage sags are known to cause frequent disruption to industrial processes and even damage machines with accumulated effect [42]. Accurate voltage sag estimation at the buses of interest, be they monitored or not, is critical, and it can be used to assist with the voltage sag mitigation planning aspects to avoid or reduce the unnecessary financial losses.

Voltage Sag Estimation (VSE) is complex, and it depends on several factors with uncertainties including system operating conditions with uncertain generation outputs from intermittent renewable energy-based generation resources, uncertainties associated with varying load demands, uncertainties associated with system faults, etc. A power system's voltage sag performance can be established reasonably accurately via long-term monitoring at enough locations within the grid and the application of statistical analysis on the monitored data. The widely used System Average RMS Variation Frequency Index (SARFI) to evaluate a system's voltage sag performance can be estimated using a filtering method [43]. Although such an approach provides a relatively accurate VSE of a system using the actual voltage sag records [44], the approach is highly reliant on the availability of historical voltage sag measurement data and it does not provide adequate details on the estimated voltage sag performance. In this case, a hybrid data-driven and model-based approach can be used in cases where the voltage sag-related historical data are not sufficient. Analysis can be implemented using a classical state estimation formulation, combined with a historical record of voltage sag measurements [45].

For the scenario mentioned above, VSE was carried out based on the measurements/data at buses where the monitors are equipped. In practice, there may be no monitored data for a majority of the buses in a practical power system. Estimating the voltage sag performances at these unmonitored buses is a more complex problem than the general VSE mentioned above. For these estimation problems, the estimation can be derived from the historical data at the monitored buses using a measurement matrix [46]. There are also other ways to achieve the goal of estimating the sag performance at unmonitored buses; these include VSE based on fault location or BDA- or AI-based methods [10].

VSE based on fault location: Events that lead to voltage sags in the network include system faults, equipment (such as transformers, reactors, etc.) energization, load energization, etc. If the voltage sag occurrence is not captured during the faults, the voltage sag performance can be obtained through equipment fault rates. As part of system reliability management, system equipment failure rates, affecting system faults, are generally recorded [47]. With this information, the voltage sag performance can be estimated using a probabilistic assessment approach, together with a Monte Carlo simulation on the given fault rates [48]. As these methods are probability-based analyses, they are not designed to locate the exact fault location [49]. However, if the locations of faults can be localized via the monitoring system, more detailed profiles of the voltage sags can be estimated. Using these approaches, the voltage sag performance at various system buses can be generated once the faults are identified [50]. These approaches are known as fault position-based sag estimation, which involves the procedure of identifying a fault location that produces the profiles (of voltages and currents) that match the measurements received from a limited set of network monitors [51]. If faults of different types are considered, each type of fault can be identified separately first, and then can be combined using an array of monitors covering different scenarios. In [52], it identified the fault areas first and later identified the faulted lines from fault simulations. Monitor Reach Area (MRA) method is widely used for voltage sag registration based on faults [53]. It defines the region of the network where the voltage sags (caused by the system faults) can be registered by the monitors. The MRA matrix is built for each type of fault, and many MRA matrices can be built to cover various fault scenarios. These approaches depend on the comparison between the actual voltages and the defined voltage thresholds. In this case, the effectiveness of the estimation is greatly dependent on the voltage thresholds. Thus, the selection of appropriate voltage thresholds is essential to ensure an accurate VSE. Otherwise, the information derived from comparing the fixed voltage thresholds may lead to misleading results or inaccuracies in the VSE. The issue is more severe with the presence of uncertainties in measurements, and measurement uncertainties exist and are prevalent in real-world systems. To resolve these issues, Ref. [10] presents a fault position-based VSE using an ANN. With a limited set of measurements, the ANN was able to accurately identify the faults. Then, based on those identified faults, the voltage sags were estimated at the unmonitored buses. This approach avoids the step of setting fixed voltage thresholds, which was used in MRA-based methods. This approach (the fault location-based estimation, together with the use of an ANN, used to estimate the fault location) is robust to uncertainties in system measurements.

VSE based on direct AI application: In the literature, usually the VSE based on a fault location (introduced earlier) requires several subprocesses, such as fault analysis (fault identification and classification) and fault simulation, followed by a VSE. To avoid these subprocesses, the hidden connection between the voltages at the monitored and unmonitored buses can be identified using various AI techniques directly, based on which, then, the voltage sag performance at the unmonitored buses can be estimated without performing simulations [54]. However, the AI model should be pretrained to generate real-time estimation of voltage sag performance. In [33], the connection between the monitored and unmonitored buses are learned using deep learning models. The models should be properly designed according to a specific problem, as the selected designs will significantly impact the VSE accuracy. Figure 5 illustrates an example deep learning model used for a VSE application. If sufficient training data were used, the

integrated deep learning models have the capability to capture the relationship between the measurement (inputs) and the voltage sag performance (outputs), and may account for any uncertainties in the network operating conditions. Once a proper learning model is trained well, the voltage sag performance results can be almost instantaneously generated when the measurement data are fed into the trained model. Different from the fault simulation-based approaches, the AI-based approaches can produce the results using a limited amount of measurements as inputs, and it can generate the results almost instantaneously without the need for other simulation or analysis steps. Another big advantage of these approaches is that the voltage sag performance can be accurately estimated regardless of the uncertain operating conditions, such as the uncertainty of renewable energy outputs, load consumption, etc.

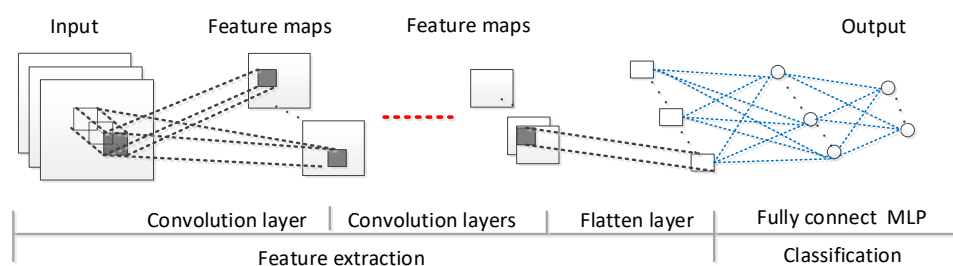


Figure 5. Illustration of the process of deep learning used for VSE.

4.4. BDA Applications in Power System Management

While BDA plays a significant role in load forecasting, fault detection, and voltage sag estimation, all meaningful processes in power system operation, there is additional potential for BDA usage in management applications. These applications utilize BDA techniques to discuss methods of energy security, infrastructure maintenance, renewable energy forecasting, etc. [55].

Stable grid operation relies not only on power quality and transmission but also on the status of the electrical equipment. In this way, BDA creates great opportunities for predictive maintenance. SCADA and advanced metering infrastructure (AMI) data record the functional status of devices and equipment so authorities can preemptively schedule maintenance and the necessary operational changes [55]. For example, BDA techniques can carry out rapid fault detection in renewable energy sources, alerting operators and allowing for more responsive repair times [56]. Predictive maintenance facilitated by BDA techniques presents opportunities for cost-effective infrastructure upgrades.

Electricity theft is considered the top cause of nontechnical energy loss [55]. Therefore, preventative measures and detection methods are of great interest to utilities and grid operators. BDA techniques can be utilized to compare standard usage patterns for customers with current measured data [55]. Trained models are able to identify energy profiles for industrial and commercial customers to differentiate between typical usage and potential theft [57]. Ref. [58] details a novel energy theft detector that predicts consumption patterns. BDA techniques are demonstrated to contribute greatly to a significant source of energy loss in power systems.

Load forecasting is often discussed in the literature and represents one part of forecasting for more effective management. Another side of forecasting is renewable energy forecasting which is essential for avoiding expensive and disruptive mismatches between generation and demand [55]. Renewable energy sources are intermittent and reliant on uncontrollable factors like weather, climate, etc. BDA techniques can identify patterns in this fluctuating data and preemptively predict renewable energy generation [59].

4.5. Future Implementation Opportunities and Barriers

BDA has the potential to transform current and future power systems by providing utilities with the insights those utilities need to empower their flexible, efficient, and

resilient digital grids, while also operating those grids cost effectively and with improved customer satisfaction. BDA demonstrates great potential in current and future applications in different areas of power systems. The key references for different aspects in this paper have been summarized and listed in Table 1.

Table 1. A list of key references in the paper.

Topics	References
Big Data Characteristics	[11–22]
BDA techniques	ANN [5,8,9,27]; Deep Learning [28,30–32]
BDA applications	Load Forecasting [8,9,38,40,41]; Fault or Outage Detection and Diagnosis [10]; Voltage Sag Estimation [43–46]; Power System Management [55–59]

In future implementation, BDA can be used to support grid management and control. BDA can be used to optimize the operation of power grids, improve energy efficiency, and reduce carbon emissions. By collecting and analyzing large amounts of data from smart meters, sensors, and other grid devices, utilities can gain insights into energy consumption patterns, identify areas of congestion or overload, and predict equipment failures before they occur. This allows utilities to take proactive measures to maintain grid stability and improve the overall reliability of their grids.

BDA can play a critical role in the integration of renewable energy sources, such as wind and solar power into the grid. By analyzing weather data and other environmental factors, utilities can predict the output of renewable energy sources and adjust their operations accordingly, a process which operates similarly to load forecasting. This can help to reduce generation curtailment and ensure a stable supply of electricity for end users.

BDA can also be used to improve customer engagement and satisfaction [60]. By analyzing customer data, such as energy usage patterns, preferences, and feedback, utilities can personalize their services and offer customized energy-saving solutions to their customers. This can help to improve customer loyalty and reduce churn.

While BDA has the potential to make great contributions to the management and operation of electrical energy systems, the collection and use of such complex data face major challenges. Data/asset management is one of the main challenges in current BDA applications [61]. Without sufficient good data (with good quality and representation considering different scenarios), the performance of BDA cannot be guaranteed. The literature identifies six areas to be addressed in the development and implementation of BDA for smart energy systems: IT infrastructure, data collection and governance, data integration and sharing, data processing and analysis, security and privacy, and professionals in BDA and smart energy management [62]. Effective data storage and computational capabilities are at the forefront of BDA technology requirements, but even as what constitutes useful data continue to grow, so too do concerns for privacy and security. Prevailing barriers to BDA implementation outside of technical requirements hinge on industry structures and organizational changes that are not designed to optimally process the influx of data sources and possible derived value that comes with BDA [63]. From both a technical and management perspective, the value created by BDA implementation is limited only by the ability to store, process, and derive meaning from disparate datasets. Apart from gathering the good quantity and quality of the training data, appropriate BDA can be designed/tailored to address specific problems in various applications.

5. Conclusions

Despite infrastructure barriers (both technological and managerial), with the continuous development of digitalization in the power/energy sector and the further improvement of BDA implementation in the real world, BDA will continue to revolutionize the way electricity is generated, distributed, and consumed. BDA for smart energy systems draws

data from a large variety of sources and has the potential to vastly improve power system operations. This paper provides a detailed overview of data sources and cutting-edge BDA research that has the potential to be used in power grid related applications. The paper introduces the techniques to integrate the BDA into a power system context and presents the capabilities of BDA in processing complex power system data to extract useful information that can be used to assist in power system operations. This paper also especially discusses typical applications of BDA in energy systems, including load forecasting, fault diagnosis, and voltage sag estimation, which are critical functions needed by the system operators to plan and operate the power system securely and optimally. The paper has also identified the potential BDA applications in power systems; in conclusion to these applications, BDA has presented its capability in supporting power system operation and planning and can be applied in the cases where unknown conditions can be estimated, forecasted, or identified based on what is known, even when the relationships between them are not straightforward. These BDA applications will pave the way for more complex applications that can contribute to the development of smart and flexible power grids. Furthermore, BDA will bring more benefits to the energy evolution, such as improved grid efficiency, active engagement from customers in supporting grid operation, and an enhanced power supply with greener energy and high-quality power.

Author Contributions: Conceptualization, H.L.; writing—original draft preparation, H.L. (Sections 3 and 4) and E.M. (Section 2); review and editing, S.C.V.; revision, E.M. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

AI	Artificial Intelligence
ANN	Artificial Neural Network
BDA	Big Data Analytics
CNN	Convolutional Neural Network
DNN	Deep Neural Networks
DSM	Demand Side Management
DT	Decision Tree
EV	Electric Vehicle
FC	Fully Connected
IT	Information Technology
KNN	K-Nearest Neighbors
LCT	Low Carbon Technology
LSTM	Long Short-Term Memory
MLP	Multilayer Perceptrons
MRA	Monitor Reach Area
PMU	Phasor Measurement Unit
RF	Random Forest
RNN	Recurrent Neural Network
SAFRI	System Average RMS Variation Frequency Index
SCADA	Supervisory Control and Data Acquisition
SVM	Support Vector Machine
V2G	Vehicle-To-Grid
VSE	Voltage Sag Estimation

References

1. Yan, Y.; Sheng, G.; Qiu, R.C.; Jiang, X. Big Data Modeling and Analysis for Power Transmission Equipment: A Novel Random Matrix Theoretical Approach. *IEEE Access* **2017**, *6*, 7148–7156. [CrossRef]
2. De Mauro, A.; Greco, M.; Grimaldi, M. A formal definition of Big Data based on its essential features. *Libr. Rev.* **2016**, *65*, 122–135. [CrossRef]
3. Zhang, Y.; Huang, T.; Bompard, E.F. Big data analytics in smart grids: A review. *Energy Inform.* **2018**, *1*, 8. [CrossRef]
4. Akhavan-Hejazi, H.; Mohsenian-Rad, H. Power systems big data analytics: An assessment of paradigm shift barriers and prospects. *Energy Rep.* **2018**, *4*, 91–100. [CrossRef]
5. Gurney, K. *An Introduction to Neural Networks*; Taylor & Francis, Inc.: London, UK, 1997.
6. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
7. BEIS; Ofgem. Transitioning to a Net Zero Energy System: Smart Systems and Flexibility Plan 2021. 2021. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1003778/smart-systems-and-flexibility-plan-2021.pdf (accessed on 16 March 2023).
8. EPRI. *Enhancements to ANNSTLF, EPRI's Short Term Load Forecaster*; Pattern Recognition Technologies, Inc.: Dallas, TX, USA, 1997.
9. Khotanzad, A.; Afkhami-Rohani, R.; Maratukulam, D. ANNSTLF-Artificial Neural Network Short-Term Load Forecaster-generation three. *IEEE Trans. Power Syst.* **1998**, *13*, 1413–1422. [CrossRef]
10. Liao, H.; Anani, N. Fault Identification-based Voltage Sag State Estimation Using Artificial Neural Network. *Energy Procedia* **2017**, *134*, 40–47. [CrossRef]
11. Angadi, R.V.; Venkataramu, P.S.; Daram, S.B. Role of Big Data Analytics in Power System Application. *E3S Web Conf.* **2020**, *184*, 01017. [CrossRef]
12. Zhu, T.; Xiao, S.; Li, Y.; Yi, P.; Gu, Y.; Zhang, Q. Emergent Technologies in Big Data Sensing: A Survey. *Int. J. Distrib. Sens. Netw.* **2015**, *11*, 902982. [CrossRef]
13. Kerai, M. *Information about the Smart Meters Statistics in Great Britain, Quarterly Report to End September 2022*; Department for Business, Energy and Industrial Strategy (BEIS): London, UK, 2022. Available online: <https://www.gov.uk/government/collecti- ons/smart-meters-statistics> (accessed on 16 March 2023).
14. Huang, Z.; Luo, H.; Skoda, D.; Zhu, T.; Gu, Y. E-Sketch: Gathering large-scale energy consumption data based on consumption patterns. In Proceedings of the 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 27–30 October 2014.
15. Li, N.; Xu, M.; Cao, W.; Gao, P. Researches on data processing and data preventing technologies in the environment of big data in power system. In Proceedings of the 2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), Changsha, China, 26–29 November 2015; pp. 2491–2494.
16. Sagioglu, S.; Terzi, R.; Canbay, Y.; Colak, I. Big data issues in smart grid systems. In Proceedings of the 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA), Birmingham, UK, 20–23 November 2016.
17. Russom, P. Big Data Analytics. In *TDWI Best Practices Report*; The Data Warehousing Institute (TDWI), 1105 Media Inc.: Renton, WA, USA, 2011.
18. Arghandeh, R.; Zhou, Y. *Big Data Application in Power Systems*; Elsevier: Amsterdam, The Netherlands, 2011.
19. Qiu, Y.; Feng, Y.; Sun, J.; Zhang, W.; Infield, D. Applying thermophysics for wind turbine drivetrain fault diagnosis using SCADA data. *IET Renew. Power Gener.* **2016**, *10*, 661–668. [CrossRef]
20. Huang, Y.; Warnier, M.; Brazier, F.M.T.; Miorandi, D. Social Networking for Smart Grid Users: A Preliminary Modeling and Simulation Study. In Proceedings of the 2015 IEEE 12th International Conference on Networking, Sensing and Control, Taipei, Taiwan, 9–11 April 2015.
21. Moreno-Munoz, A.; Bellido-Outeirino, F.; Siano, P.; Gómez-Nieto, M. Mobile social media for smart grids customer engagement: Emerging trends and challenges. *Renew. Sustain. Energy Rev.* **2016**, *53*, 1611–1616. [CrossRef]
22. Alghamdi, T.A.; Javaid, N. A Survey of Preprocessing Methods Used for Analysis of Big Data Originated from Smart Grids. *IEEE Access* **2022**, *10*, 29149–29171. [CrossRef]
23. Wang, H.; Yemeni, Z.; Ismael, W.M.; Hawbani, A.; Alsamhi, S.H. A reliable and energy efficient dual prediction data reduction approach for WSNs based on Kalman filter. *IET Commun.* **2021**, *15*, 2285–2299. [CrossRef]
24. Koprinska, I.; Rana, M.; Agelidis, V.G. Correlation and instance based feature selection for electricity load forecasting. *Knowl. Based Syst.* **2015**, *82*, 29–40. [CrossRef]
25. Saleh, A.I.; Rabie, A.H.; Abo-Al-Ez, K.M. A data mining based load forecasting strategy for smart electrical grids. *Adv. Eng. Inform.* **2016**, *30*, 422–448. [CrossRef]
26. Li, Z.; Liu, J.; Lin, Y.; Wang, F. Grid-Constrained Data Cleansing Method for Enhanced Bus Load Forecasting. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [CrossRef]
27. Haque, M.T.; Kashtiban, A.M. Application of Neural Networks in Power Systems; A Review. *Int. J. Energy Power Eng.* **2007**, *1*, 897–901.
28. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]

29. Brahma, P.P.; Wu, D.; She, Y. Why Deep Learning Works: A Manifold Disentanglement Perspective. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *27*, 1997–2008. [CrossRef]
30. Nielson, M.A. *Neural Networks and Deep Learning*; Determination Press, 2015; Available online: <http://neuralnetworksanddeeplearning.com/> (accessed on 16 March 2023).
31. Varga, E.D.; Beretka, S.F.; Noce, C.; Sapienza, G. Robust Real-Time Load Profile Encoding and Classification Framework for Efficient Power Systems Operation. *IEEE Trans. Power Syst.* **2014**, *30*, 1897–1904. [CrossRef]
32. Claessens, B.J.; Vranx, P.; Ruelens, F. Convolutional neural networks for automatic state-time feature extraction in reinforcement learning applied to residential load Control. *IEEE Trans. Smart Grid* **2016**, *9*, 3259–3269. [CrossRef]
33. Liao, H.; Milanović, J.V.; Rodrigues, M.; Shenfield, A. Voltage Sag Estimation in Sparsely Monitored Power Systems Based on Deep Learning and System Area Mapping. *IEEE Trans. Power Deliv.* **2018**, *33*, 3162–3172. [CrossRef]
34. Mazhar, T.; Asif, R.N.; Malik, M.A.; Nadeem, M.A.; Haq, I.; Iqbal, M.; Kamran, M.; Ashraf, S. Electric Vehicle Charging System in the Smart Grid Using Different Machine Learning Methods. *Sustainability* **2023**, *15*, 2603. [CrossRef]
35. Barakat, E.; Qayyum, M.; Hamed, M.; Al Rashed, S. Short-term peak demand forecasting in fast developing utility with inherit dynamic load characteristics. I. Application of classical time-series methods. II. Improved modelling of system dynamic load characteristics. *IEEE Trans. Power Syst.* **1990**, *5*, 813–824. [CrossRef] [PubMed]
36. Fidalgo, J.; Lopes, J. Forecasting active and reactive power at substations transformers. In Proceedings of the 2003 IEEE Bologna Power Tech Conference Proceedings, Bologna, Italy, 23–26 June 2003.
37. Bhatt, A.K.; Solanki, P.; Bhatt, A.; Cherukuri, R. A fast and efficient back propagation algorithm to forecast active and reactive power drawn by various capacity Induction Motors. In Proceedings of the International Conference on Circuits, Power and Computing Technologies (ICCPCT), Nagercoil, India, 20–21 March 2013.
38. Khotanzad, A.; Afkhami-Rohani, R.; Lu, T.-L.; Abaye, A.; Davis, M.; Maratukulam, D. ANNSTLF—a neural-network-based electric load forecasting system. *IEEE Trans. Neural Netw.* **1997**, *8*, 835–846. [CrossRef]
39. Park, D.; El-Sharkawi, M.; Marks, R.; Atlas, L.; Damborg, M. Electric load forecasting using an artificial neural network. *IEEE Trans. Power Syst.* **1991**, *6*, 442–449. [CrossRef]
40. Oonsivilai, A.; El-Hawary, M.E. Wavelet neural network based short term load forecasting of electric power system commercial load. In Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering, Edmonton, AB, Canada, 9–12 May 1999.
41. Taylor, J.; Buizza, R. Neural network load forecasting with weather ensemble predictions. *IEEE Trans. Power Syst.* **2002**, *17*, 626–632. [CrossRef]
42. Chan, J.Y.; Milanovic, J.V.; Delahunty, A. Risk-Based Assessment of Financial Losses Due to Voltage Sag. *IEEE Trans. Power Deliv.* **2011**, *26*, 492–500. [CrossRef]
43. Zambrano, X.; Hernandez, A.; Izzeddine, M.; de Castro, R.M. Estimation of Voltage Sags from a Limited Set of Monitors in Power Systems. *IEEE Trans. Power Deliv.* **2017**, *32*, 656–665. [CrossRef]
44. Bollen, M.H.J. *Understanding Power Quality Problems: Voltage Sags and Interruptions*; Wiley: New York, NY, USA, 2000.
45. Short, T.; Mansoor, A.; Sunderman, W.; Sundaram, A. Site variation and prediction of power quality. *IEEE Trans. Power Deliv.* **2003**, *18*, 1369–1375. [CrossRef]
46. Espinosa-Juarez, E.; Hernández, A. A Method for Voltage Sag State Estimation in Power Systems. *IEEE Trans. Power Deliv.* **2007**, *22*, 2517–2526. [CrossRef]
47. Brown, R.E. *Electric Power Distribution Reliability (Power Engineering (Willis))*, 2nd ed.; Willis, H.L., Ed.; CRC Press: Boca Raton, FL, USA, 2008.
48. Li, W. *Risk Assessment of Power Systems: Models, Methods, and Applications*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014.
49. Myo Thu, A.; Milanovic, J.V. Stochastic prediction of voltage sags by considering the probability of the failure of the protection system. *IEEE Trans. Power Deliv.* **2006**, *21*, 322–329.
50. *IEEE Std 493-2007 (Revision of IEEE Std 493-1997)*; IEEE Recommended Practice for the Design of Reliable Industrial and Commercial Power Systems—Redline. The Institute of Electrical and Electronics Engineers: New York, NY, USA, 2007; pp. 1–426.
51. Dugan, R.C.; McGranaghan, M.; Santoso, S.; Beaty, H.W. *Electrical Power Systems Quality*; McGraw-Hill: Berkshire, UK, 2003.
52. Majidi, M.; Etezadi-Amoli, M.; Fadali, M.S. A sparse-data-driven approach for fault location in transmission networks. *IEEE Trans. Smart Grid* **2017**, *8*, 548–556. [CrossRef]
53. Olguin, G.; Vuinovich, F.; Bollen, M. An Optimal Monitoring Program for Obtaining Voltage Sag System Indexes. *IEEE Trans. Power Syst.* **2006**, *21*, 378–384. [CrossRef]
54. Espinosa-Juarez, E.; Hernandez, A. Neural Networks Applied to Solve the Voltage Sag State Estimation Problem: An Approach Based on the Fault Positions Concept. In Proceedings of the 2009 Electronics, Robotics and Automotive Mechanics Conference (CERMA), Washington, DC, USA, 22–25 September 2009.
55. Dhupia, B.; Rani, M.U.; Alameen, A. The Role of Big Data Analytics in Smart Grid Management. In *Emerging Research in Data Engineering Systems and Computer Communications*; Venkata Krishna, P., Obaidat, M., Eds.; Advances in Intelligent Systems and Computing; Springer: Singapore, 2020; Volume 1054, pp. 403–412.
56. Ma, Z.; Xie, J.; Li, H.; Sun, Q.; Si, Z.; Zhang, J.; Guo, J. The Role of Data Analysis in the Development of Intelligent Energy Networks. *IEEE Netw.* **2017**, *31*, 88–95. [CrossRef]

57. Ahmad, T.; Chen, H.; Wang, J.; Guo, Y. Review of various modeling techniques for the detection of electricity theft in smart grid environment. *Renew. Sustain. Energy Rev.* **2018**, *82*, 2916–2933. [[CrossRef](#)]
58. Jokar, P.; Arianpoo, N.; Leung, V.C.M. Electricity Theft Detection in AMI Using Customers' Consumption Patterns. *IEEE Trans. Smart Grid* **2015**, *7*, 216–226. [[CrossRef](#)]
59. Landa-Torres, I.; Unanue, I.; Angulo, I.; Russo, M.R.; Campolongo, C.; Maffei, A.; Srinivasan, S.; Glielmo, L.; Iannelli, L. The application of the data mining in the integration of RES in the smart grid: Consumption and generation forecast in the I3RES project. In Proceedings of the 2015 IEEE 5th International Conference on Power Engineering, Energy and Electrical Drives (POWERENG), Riga, Latvia, 11–13 May 2015.
60. Mathumitha, R.; Rathika, P.; Manimala, K. Big Data Analytics and Visualization of Residential Electricity Consumption Behavior based on Smart Meter Data. In Proceedings of the 2022 International Conference on Breakthrough in Heuristics and Reciprocation of Advanced Technologies (BHARAT), Visakhapa, India, 7–8 April 2022.
61. Koziel, S.; Hilber, P.; Ichise, R. Application of big data analytics to support power networks and their transition towards smart grids. In Proceedings of the 2019 IEEE International Conference on Big Data, Los Angeles, CA, USA, 9–12 December 2019; pp. 6104–6106.
62. Zhou, K.; Fu, C.; Yang, S. Big data driven smart energy management: From big data to big insights. *Renew. Sustain. Energy Rev.* **2016**, *56*, 215–225. [[CrossRef](#)]
63. Wamba, S.F.; Akter, S.; Edwards, A.; Chopin, G.; Gnanzou, D. How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *Int. J. Prod. Econ.* **2015**, *165*, 234–246. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.