SPRINGER BRIEFS IN COMPUTER SCIENCE

Murad Khan
Bilal Jan
Haleem Farman

Deep Learning: Convergence to Big Data Analytics

Springer

SpringerBriefs in Computer Science

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

Briefs allow authors to present their ideas and readers to absorb them with minimal time investment. Briefs will be published as part of Springer's eBook collection, with millions of users worldwide. In addition, Briefs will be available for individual print and electronic purchase. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, easy-to-use manuscript preparation and formatting guidelines, and expedited production schedules. We aim for publication 8–12 weeks after acceptance. Both solicited and unsolicited manuscripts are considered for publication in this series.

More information about this series at http://www.springer.com/series/10028

Murad Khan · Bilal Jan · Haleem Farman

Deep Learning: Convergence to Big Data Analytics

Springer

Murad Khan
Department of Computer Science
Sarhad University of Science
and Information Technology
Peshawar, Pakistan

Haleem Farman
Department of Computer Science
Islamia College Peshawar
Peshawar, Pakistan

Bilal Jan
Department of Computer Science
Fata University
FR Kohat, Pakistan

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-981-13-3458-0 ISBN 978-981-13-3459-7 (eBook)
https://doi.org/10.1007/978-981-13-3459-7

Library of Congress Control Number: 2018962781

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

About This Book

This book presents deep learning techniques, concepts, and algorithms to classify and analyze big data. The big data analytics using traditional techniques have various challenges and issues such as high speed analytics, accuracy, and efficient processing of big data in real time. In addition, the Internet of things is progressively increasing in various fields such as smart cities, smart homes, and e-health. This enormous number of connected devices generate a huge amount of data every day, and hence, we need sophisticated algorithms to deal, organize, and classify such huge amount of data in less processing time and space. Similarly, the existing techniques and algorithms for deep learning in big data field have several advantages because of the two main branches of deep learning, i.e., convolution and deep belief networks. This book gives an insight into all those techniques and applications based on these two types of deep learning. Also, the book helps the students, researchers, and newcomers in understanding big data analytics based on deep learning approaches. Further, the book gives an introductory-level understanding to the new programming languages and tools used to analyze big data specifically in real time such as Hadoop, Spark, and GraphX. Various machine learning techniques are discussed in concatenation with the deep learning paradigm to support high-end data processing, data classifications, and real-time data processing issues.

The classification and presentation of the book are kept quite simple to help the readers and students understand the basic concepts of various deep learning paradigms and frameworks. The book mainly focuses on theory instead of presenting the mathematical background of the deep learning concepts, because the theory behind these techniques still needs further improvements and refining to enable the students and practitioners to understand and develop their mathematical models. The book consists of five chapters beginning with an introductory explanation of the big data alongside deep learning techniques, followed by integration of big data and deep learning techniques, and finally the future directions in big data analytics using deep learning techniques.

Aims and Scope of the Book

Deep learning techniques are widely adopted in many fields of computer science and engineering for data analysis. The introduction of big data analytics has several challenges such as processing of huge amount of data efficiently and in less time. However, the conventional methods and database management systems fail to offer such functionalities of processing huge data in less and finite time. Therefore, the researchers try various techniques and algorithms from various fields to come up with a system which can process data as quick as possible. The Hadoop ecosystem is developed for such problems with a map and reduce concept. The Hadoop ecosystem can process a huge amount of data; however, it cannot offer the functionality of processing data in real time. Therefore, sophisticated techniques are need of the day to deal with such huge amount of data generated in real time with high velocity, volume, and varsity (i.e., 3V concept). In addition, the big data is always generated from heterogeneous sources; therefore, learning functionalities with conventional SQL query techniques always require high-end nodes. However, installing high-end node everywhere is a difficult and impossible task. Thus, incorporated learning technique without installing high-end node can be possible using deep learning techniques. The aim of this book is to bring the academia, research community, and industry at one platform to design and suggest deep learning techniques for big data analytics. The classification of the book is further divided into four main categories: (1) novel and sophisticated real-time deep learning analytics for cross-layer and inter-domain services, (2) state-of-the-art theories and practice for application scenarios from heterogeneous and cross-domain big data analytics, (3) survey and tutorials from world's top research community and laborites working in big data analytics, and (4) building of benchmark and test beds for analyzing big data using deep learning techniques. Further, the scope of the book consists of the following topics:

- Feature and data-pattern learning using deep learning techniques
- Large-scale data processing-based learning techniques
- Architectural design of deep learning algorithms and techniques
- Pattern recognition via deep learning techniques

- Innovative scientific methodologies for big data processing in real time
- Big data processing for urban planning, management, and sustainability
- Sustainability of business models for smart home, smart cities, e-health care
- The analysis and processing of big data generated from Internet of things
- Information privacy and data leakage prevention using big data analytics
- Novel and efficient machine and data mining on big data
- Big data sets and benchmarks generation and processing for big data analytics based on deep learning techniques.

Contents

About the Authors

Murad Khan received his BS in computer science from the University of Peshawar, Pakistan, in 2008. He completed his Ph.D. in computer science and engineering from School of Computer Science and Engineering in Kyungpook National University, Daegu, Korea. He has published over 60 international conference and journal papers along with two chapters in Springer and CRC Press. He also served as a TPC member in world-reputed conferences and as a reviewer in numerous journals such as Future Generation Systems (Elsevier) and IEEE Access. In 2016, he was awarded Qualcomm Innovation Award at Kyungpook National University for designing a smart home control system. He was also awarded Bronze Medal in ACM SAC 2015, Salamanca, Spain, on his distinguished work in multi-criteria-based handover techniques. He is a member of various communities such as ACM and IEEE, and CRC Press. His areas of expertise include ad hoc and wireless networks, architecture designing for Internet of things, communication protocols designing for smart cities and homes, big data analytics, etc.

Bilal Jan received his MS and Ph.D. from the Department of Control and Computer Engineering (DAUIN), Politecnico di Torino, Italy, in 2010 and 2015, respectively. He has published several papers in reputed journals and conferences. He is currently working as Assistant Professor and Head of the Department of Computer Science, FATA University, Darra Adam Khel, FR Kohat, Pakistan. He is a reviewer for numerous leading journals. His research interests include general-purpose programming in GPUs, high-performance computing, wireless sensor networks, Internet of things (IoT), deep learning, and big data.

Haleem Farman received his MS from the International Islamic University, Islamabad, Pakistan, in 2008. He is currently pursuing his Ph.D. in Computer Science in the Department of Computer Science, University of Peshawar, Pakistan, and working as Lecturer in the Department of Computer Science, Islamia College

Peshawar, Pakistan. He has authored/co-authored more than 20 research papers in respected journals and conferences. In addition, he serves as an invited reviewer for several journals, such as Elsevier Sustainable Cities and Society. His fields of interest include wireless sensor networks, Internet of things, big data analytics, privacy, optimization techniques, and quality of service issues in wireless networks.

List of Figures

Chapter 1
Introduction

Bilal Jan, Haleem Farman and Murad Khan

Abstract Recently, deep learning techniques are widely adopted for big data analytics. The concept of deep learning is favorable in the big data analytics due to its efficient use for processing huge and enormous data in real time. This chapter gives a brief introduction of machine learning concepts and its use in the big data. Similarly, various subsections of machine learning are also discussed to support a coherent study of the big data analytics. A thorough study of the big data analytics and the tools required to process the big data is also presented with reference to some existing and well-known work. Further, the chapter is concluded by connecting the deep learning with big data analytics for filling the gap of using machine learning for huge datasets.

List of Acronyms

AI	Artificial intelligence
ANN	Artificial neural networks
HDFS	Hadoop Distributed File System
M2M	Machine to machine
IoT	Internet of things
CPS	Cyber physical systems
ICN	Information-centric networking
WSN	Wireless sensor network

1.1 Machine Learning

The need of machine learning was felt when artificial intelligence (AI)-based systems were facing difficulties with hard-coded programs, and it was suggested that machines should be able to extract patterns from the data by itself without the involvement of human or programs for specific tasks. The idea of machine learning was

M. Khan et al., *Deep Learning: Convergence to Big Data Analytics*, SpringerBriefs in Computer Science, https://doi.org/10.1007/978-981-13-3459-7_1

introduced in 1959 by Arthur Samuel (field expert) that instead of programming machines for specific tasks, computers should be able to learn themselves (BMC blog 2018). Machine learning is the subset of artificial intelligence, in which system can adjust its activities and react to specific situation when provided with large amount of data (Ahmad et al. 2018a, b). In machine learning, systems are trained to act accordingly. These systems are provided with many examples specifically related to task, and statistical structures are identified that leads system to define rules for that particular task. Machine learning deals with large amount of datasets (Khumoyun et al. 2016), for instance, medical dataset containing millions of patient images for different diseases. There are many applications such as recommendation and navigation systems using machine learning giving more accurate and efficient result as compared to hard-coded programs. Machine learning is classified into supervised learning and unsupervised learning (BMC blog 2018).

1.1.1 Supervised Learning

In supervised learning, the machine is trained according to the given input data and output is drawn from it accordingly. The relation between input data and expected output is approximated through machine learning algorithms. In simple words, supervised learning is based on the prior knowledge that what our output will be. Normally supervised learning is done in two ways, either as classification or regression. In classification, the input is mapped to output tags, while in regression the input is mapped to a continuous output. The goal of both ways is to find structures in input data that can be transformed into effective, accurate, and correct output. In real world, it is not always the case that the data labels will always be correct. Incorrect data labels can lead to ineffective output and will clearly affect the effectiveness of the model. Most common algorithms are naive bayes, artificial neural networks, support vector machines, and logistic regression (Towards Data Science 2018).

1.1.2 Unsupervised Learning

Unsupervised learning works on input data only by not considering the labeled output. It finds structures in the data without having expected output. Unsupervised learning algorithms are complex and time-consuming because of generating output without prior knowledge. Clustering, dimensionality reduction, and representation learning are the most commonly used tasks in unsupervised learning. The most common algorithms used in unsupervised learning are K-means clustering, autoencoder, and principal component analysis. Dimensionality reduction and exploratory analysis are the two most common use cases for unsupervised learning. In dimensionality analysis, feature or columns are reduced to represent data, while in exploratory analysis, structure in data can be automatically identified using unsupervised learning.

The trend toward machine learning is getting bigger and bigger as it is assisting business a great deal such as machine learning in healthcare, financial services, vehicle automation, and retails. Data in digital form is generated in huge volume and variety. It is not important to gather more data, but how effectively and efficiently can we use this data is the problem to be solved. Machine learning systems are really helping companies to manage, analyze, and produce output from such huge and diverse data. Machine learning systems can identify the hidden patterns in data, and customer preferences can be identified to boost business, market trends, and many more ways in which business can be boost. As the data is growing data by data, companies such as Facebook and Google are more interested in customer behavior to improve their service. To handle big data, a subset of machine learning known as deep learning is used, in order to use the available data in more effective and efficient manner.

1.2 Deep Learning

Deep learning uses artificial neural networks (ANN) that got inspiration from the neuron present in human brain. It consists of layers, and the word deep refers to the depth of layers (Ahmad et al. 2018a, b). Initially, the word deep referred to very few layers, but due to the use deep learning in complex problems, the number of layers is in hundreds and even thousands. Deep learning is very successful in many domains such as image processing, healthcare, transportation, and agriculture; with the help of deep learning, more and more data can be utilized in best possible and it is getting popularity due to the availability of trained dataset, for instance, ImageNet (Deng et al. 2009) that contains thousands of images. Secondly, the low-cost GPUs are in more use to train data and can avail the services of cloud as well. Giant companies such as Facebook, Amazon, Google, and Microsoft are using deep learning techniques to analyze the huge amount of data on daily basis.

Deep learning got high attention not only from researcher but from techno-companies as well. Social media companies such as Facebook, Twitter, and YouTube generate large volume of data on daily basis due to the number of users they have (Jan et al. 2017). It is very important for them to handle this huge data often termed as “Big Data.” By using traditional data analysis tools, it is very difficult and almost impossible to have a good insight of the data and extract the meaningful data from it. Even machine learning technique will not work that much efficient as required; therefore, deep learning is used in order to analyze deep in the network to extract meaningful, accurate, and precise information. In the coming chapter, we will be studying deep learning in detail.

1.3 Conventional Data Processing Techniques

Data processing is a sequence of steps for transforming raw data into meaningful information or insights as an output in the form of plaintext file, table/spreadsheet, charts and graphs, maps/vector, or image file leading to the solution of a problem or improving an existing situation. It leads to solving a problem or improving an existing situation. Data collection, preparation as suitable input, processing, output and interpretation and storage are some core steps of data processing. It is important to process data for businesses and scientific operations where large volumes of output is required out of business data after repeated processing while scientific data requires fast-generated outputs governing numerous computations. There are many traditional data processing techniques briefly discussed as follows (BMC blog 2018; Ahmad et al. 2018a, b; Khumoyun et al. 2016; Towards Data Science 2018):

Manual Data Processing: In this method of data processing, all logical operations, calculations, data transfer from one place to another, and required results are obtained without intervention of any tool or machine. Manual data processing is performed in small firms and government institutes for their tasks manually but is avoided to be used due to error-pruned, time-consuming, and labor-intensive nature of this processing. The absence of technology or its high cost favored the use of manual data processing, but advancement of technology has drastically decreased the dependency on manual techniques.

Mechanical Data Processing: This method is more accurate and faster than manual data processing by using devices such as mechanical printers, typewriters, or other mechanical devices to work for printing press or examination boards. More sophisticated and better computing powered machines surpassed them.

Electronic Data Processing: It is a modern data processing technique, very fast and accurate, with the involvement of a computer that automatically processes the data according to the provided set of instructions along with it. Computer is another name of electronic data processing machine. Batch processing, online processing, real-time processing, multiprocessing, and time sharing and distributed processing are some methods of data processing by electronic means. Some methods are briefly discussed as follows:

- Batch Processing: It is the simplest form of data processing which works for large volume of data if data could be clumped into one or two categories; e.g., daily or weekly transactions of a store can batch-processed, and results are sent to the head office.
- Online Processing: In this method, directly attached equipment to a computer and Internet connections are utilized to allow data to be stored at one place and to get it used for processing at different places. This type of processing is used in cloud computing.
- Real-Time Processing: For an instant turnaround, this method is faster than batch processing; e.g., instant update of record is required in case of canceling a reservation or buying an airline ticket.

- Distributed Processing: It is widely used in a scenario where one server is connected with remotely located workstations, e.g., ATMs where fixed software is located at a particular place while all the end machines make use of that same set of instructions and information.
- Data Mining: Data mining collects data from different sources and combines it to analyze to find patterns, to group similar observations, to predict class label of previously unknown instances based on historical data, and to identify deviated values from the normal ones.
- Neural networks, nonlinear data modeling tools, are automated to store, identify, and find patterns in database or to decipher complex relationships between inputs and outputs.

1.4 Data Mining Techniques: A Big Data Analysis Approach

Availability of gigantic amount of data due to enormous expansion of information technology arises the strong demand for data mining techniques which play a vital role in the analysis of data using techniques of statistics, artificial intelligence, database system, and machine learning. Data mining is a computing process to explore large datasets to discover interesting patterns. It has mainly three steps (Towards Data Science 2018; Jan et al. 2017; Five Data 2018):

Walking Through: Nature of the data is determined, and it is converted into a suitable form after going through all the collected data. Educated guesses and different types of imputation method such as regression substitution, average imputation, and multiple imputation are used to substitute values to replace missing data.

Identifying Patterns: By understanding what information to extract, patterns are checked that could lead to make predictions.

Outcome Planning: Here, desired outcome is planning having patterns on hand.

On completion of data mining process, one can utilize it in a number of applications such as to recommend new products based on the patterns discovered within a data and helps to identify clusters showing the similarities and dissimilarities within the data and classifying the data to draw useful insights from it.

There are a number of data mining techniques for achieving business goals (Ahmad et al. 2018a, b).

Association Rule Learning: Here, the frequently occurring items or variables and association or relationship of two or more items are analyzed in order to uncover the hidden patterns in the large datasets. Various applications can be drawn utilizing this technique; e.g., the trend of buying bananas along with cereal in customers' purchases record can help to make better decision of placing bananas closer to cereals in the store shelf to boost selling these items together. Also, effect of coupons or sales offer on selling can be tracked.

Clustering Analysis: Similar objects exhibiting same behaviors are grouped and analyzed with this old yet popular unsupervised technique of data mining. K-means clustering, agglomerative clustering, and hierarchical clustering are some well-known clustering approaches.

Classification Analysis: It is supervised way of classifying large sets of data by predicting a class label to unknown instances based on the values in the historical database. In other words, it helps to find the categories the data belongs to. The classification methods make use of decision tree and neural network techniques. Models are built by dividing the data into two datasets—one for training or building the model and another to test the built models which are then compared with predicted values in order to be evaluated. For example, classification analysis is performed by our email provider to classify our incoming email as legitimate or spam based on some specific words of the email or its attachment.

Regression Analysis: It explores the dependency between variables b assuming a one-way causal effect of a variable in response of another. In contrast to correlation, here dependency is not necessarily from both sides; one variable can be dependent on other but not vice versa.

Anomaly or Outlier Detection: Various techniques are governed to detect, analyze, and understand anomaly, outlier, or an exception in the data that deviates from a particular dataset.

Data mining evolving nature of data has prompted the need to develop new analytic techniques by expanding already existing data mining techniques (Khumoyun et al. 2016). Social media data in the form of comments, likes, browsing behaviors, tweets, opinions, e-commerce, and medical data is some digital sources of the big data. Significant contribution of numerous sensors also played a vital role in the growth of big data (Jan et al. 2017).

Life cycle of big data analytics consists of various steps such as problem definition, problem tackling based using existing techniques, Human Resource requirement estimation, data acquisition, munging and storage, then performing exploratory analysis of data in order to select decision model and finally converting the designed and evaluated model into implementation (Deng et al. 2009; Data Processing and Data Processing Method 2018).

Traditional database management tools are not capable to store and process tremendous volume of big data (Jan et al. 2017). Hadoop and Mahout are popular big data analytics tools.

Apache Hadoop is a framework based on MapReduce programming model that provides the solution by supporting distributed data processing. Large datasets are processed in parallel across clusters of computers. Hadoop Distributed File System (HDFS) and MapReduce are two main parts of Hadoop where former is a distributed file system consisting thousands of nodes used for big data storage primarily, while latter is software that processes these nodes in parallel or simultaneously and retrieves data at a faster rate. Classification algorithms are implemented in Apache Mahout, which are applied against the MapReduce paradigm to estimate the values of the subjects under consideration.

MapReduce is a programming model encompassing map and reduce as two operations defined by the users to generate and process massive datasets stored in HDFS by allowing parallel large computations boosting the speed and reliability. Divide and conquer is the basic technique behind it where smaller chunks of the given big data problem are made, shuffled, and reduced to acquire the desired result as output. More specifically, all the complex business rules, logic, or costly code are specified first in the Map phase; then, lightweight processing, e.g., summation or aggregation, is specified in the second Reduce phase. There are multiple phases of data processing with different components by Hadoop MapReduce discussed as follows (Jan et al. 2017; Explain three methods 2018).

Step 1. A Master process is forked by the user and processed by a number of workers.

Step 2. The Map and Reduce tasks are assigned to the worker processes by the Master process.

Step 3. The input file is split into chunks of 16–64 MB by the user program of the MapReduce library. One or more input file(s) chunks are assigned to each Map task.

Step 4. Then, the data chunks are turned into a sequence of key–value pairs by these Map tasks for reading by the mapper.

Step 5. Mapper produces intermediate key–value pairs from the input record produced in the previous phase. The intermediate output is totally different from the input pair and is forwarded to the combiner for next processing.

Step 6. Combiner acts as mini-reducer to aggregate mapper's output locally by minimizing the data transfer between reducer, and mapper then forwards the output to the next practitioner step.

Step 7. Here, partitioning is performed on the output of the combiner on the basis of the key in MapReduce to evenly distribute the map output over the reducer.

Step 8. Physical movement of the data, also known as shuffling, over the network to the reducer nodes is done at this stage, and intermediate output is merged and sorted and then is shipped as input to the reduce stage.

Step 9. In this phase, intermediate key–value pairs generated by the mappers are considered as the input by the reducer and reducer function is applied on each of them to produce the final output stored on HDFS.

Step 10. In the last stages, RecordWriter writes the produced output key–value pairs from the reducer phase to the output files. Output format specified the method of writing these output key–value pairs on HDFS.

Virtual machine is a big data processor platform for analysis and processing the data (Data Processing 2018). It is a platform for machine learning and analytics optimized for the cloud.

1.5 Big Data Analytics

During the past 20 years, data has been enormously generated due to the interconnecting devices in various forms such as machine to machine (M2M), Internet of things (IoT), cyber physical systems (CPS), and information-centric networking (ICN).

However, the very formal definition was first introduced by D. Laney, META Group in 2001, by defining the three Vs architecture of big data as volume, velocity, and variety (Laney 2001). For a decade, the researchers follow the three Vs architecture; however, it was changed after 2010 when Dave Beulke and Associates introduces the five Vs architecture such as volume, velocity, value, veracity, and variety in 2011 (Associates 2011). The definition of big data consists of 6 Vs architecture that is put forward by Enterprise Strategy Group in 2012 (Group 2012). Similarly, the concept of 8 Vs, i.e., volume, velocity, value, variability, veracity, viscosity, virality, and variety, is introduced by Vorhies (2014). However, all types of data do not follow any of the aforementioned architecture as standard. It entirely depends on the researchers, scientists, and academia, who are extracting data for various purposes. Similarly, big data is not only used to process data and come up with necessary results. It is also used to find the correct tools and identify the applications for big data in smart environments such as smart homes, smart cities, smart e-health systems. Big data is defined for huge amount of data, and thus, conventional data tools and processing applications are not sufficient to analyze the big data. The market of big data involves in three layers of processing, i.e., 1) the infrastructure layer (particularly related to hardware and physical equipment), 2) the data management, organization, and analytics layer, and 3) the services layer. The infrastructure layer consists of hardware components such as cloud facilities, networking services, communication technologies. The gathering, collection, and acquisition of data normally happen at the infrastructure layer. It is sometime done with the help of sensors which further needs an optimized and efficient wireless sensor network (WSN). After the data is collected at the infrastructure layer, the data is communicated with the management layer with the help of underlying technologies. The management and analytics layer processes the data for various activities such as storing, decision-making, pattern recognition. Finally, the data is disseminated among the users with the help of services layer. The services layer also offers other functionalities such as defining the applications and interfaces where user can connect to the system to retrieve the required information. Figure 1.1 shows the functionalities and concepts provided by all the three layers.

There are various challenges involved in big data analytics such as processing data in real time, disseminating data with users, and providing them interfaces. Similarly, most of the recent literature used big data for different applications such as energy management in smart homes, cities, and buildings and urban planning. However, that is no generic standard available for analyzing big data apart from above three layers. There is need of standard solutions and regulation for analyzing and processing big data. The world famous firms such as ACM and IEEE are working hard to come up with common and generic standard for big data. However, at this stage, we will appreciate the efforts of researchers, scientist, and academia in this regard. Further, the big data is a part of data science and we cannot say that one particular algorithm can be used to process big data. Big data is a linear process, and at each step, one can use a particular type of algorithm to solve a problem. With the passage of time, various algorithms and methods have been employed by the researchers to make the process of analyzing big data more efficient and convenient for the users and

Fig. 1.1 Three layers of big data

practitioners (Khan et al. 2018a, b). The processing of data in real time required high computing resources and efficient algorithms such as deep learning, machine learning, and transfer learning. However, implementing these techniques in real time is a challenging job and the researchers are looking for various computing mechanisms such as artificial intelligence and quantum computing. As the computing system is advanced every day, there is a possibility in future to reach to certain level of processing data in real time. For example, if a space shuttle or space vehicles sent to Mars for collecting data and processing it in real time would require high-end programming and pattern matching techniques. On top of the computation capabilities, the data analytics are quite important which is however impossible with conventional data analytics. Similarly, the conventional MapReduce architecture used in the Hadoop ecosystem is mainly used to process off-line data, which is therefore nearly impossible to be used for processing real-time data. The Hadoop ecosystem alongside Spark somehow addresses the processing of data in real time. However, when it comes to huge amount of data specifically in petabyte, the performance of the Spark also degraded exponentially. Therefore, a cluster system combined of Spark can be used to process the data in petabytes (Silva et al. 2017). Similarly, integrating data sources for solving a problem or even answering a question requires big data analytics. For example, understanding the pattern of why the rainfall rate is dropping lower than the average rainfall every year may require obtaining the data from satellites, sensors, and airborne and fusing them together for a better approach. Similarly, obtaining the data of vehicles moving on some particular roads and combining them for analytics might resulting various patterns such as why the number of cars are high on a particular time of the day and so on. All these challenges discussed above can be answered with proper data analytic techniques, tools, and methods.

1.6 Deep Learning in Big Data Analytics

Machine learning is widely studied recently for processing big data in real time. Similarly, the machine learning techniques can be considered an essential part of the overall process of big data analytics. In such process, a computer system is trained with a pattern and then the computer system finds the same pattern or similar pattern in the entire data. This particular type of learning where the computer system is taught at the beginning is called supervised learning. However, training a computer system in real time is challenging task to accomplish with even hundreds and thousands of examples. On the hand in an unsupervised learning methodology, a process is applied to classify and manage the information in unstructured documents. Similarly, an analysis can be carried out to find natural divisions and clustering in the data. Further, the unsupervised learning looks for inherent sequence and patterns in the data and groups them together. Furthermore, in such processes, the outliers can be identified for grouping the data falls outside of a particular sequence or pattern. Similarly, the unsupervised learning can be applied to the existing data to find real-time solutions to disasters, heavy rainfall, etc. A number of approaches have been proposed to use machine learning for big datasets alongside Hadoop and MapReduce architectures (Dean and Ghemawat 2004; Shvachko et al. 2010). Similarly, the online and deep learning is also adopted alongside machine learning to overcome the challenges of big data. Summarizing the challenges present with big data using machine and deep learning, we come across unstructured data obtained from heterogeneous sources, high and fast streaming data, noisy and poor-quality data, high-dimensional data, and data with limited labels (Najafabadi et al. 2015; Sukumar 2014). Before applying machine learning techniques to big data analytics, one can have enough knowledge of statistical methods and signal processing techniques. The signal processing techniques play an important role in classifying data obtained from heterogeneous sources and inputs. Qiu et al. presented a work with identifying signal processing techniques to address various challenges such as large-scale, different data types, high-speed data, incomplete and uncertain data, and density of data (Qiu et al. 2016). However, dealing with large-scale data using machine learning requires high memory. Therefore, efficient systems are needed to overcome the issues of high memory requirements. A review of such systems is presented by Al-Jarrah et al. for processing large-scale data with less memory requirements (Al-Jarrah et al. 2015). The authors were mainly interested in the analytical aspects of processing big data. However, they did not address the computation complexity which is one of the main components of high memory requirements. Therefore, a more comprehensive study is still needed to properly address the challenges present in the area of using machine learning for big data analytics.

References

Ahmad J, Muhammad K et al (2018a) Visual features based boosted classification of weeds for real-time selective herbicide sprayer systems. Comput Ind 98:23–33

Ahmad J, Muhammad K et al (2018b) Efficient conversion of deep features to compact binary codes using fourier decomposition for multimedia big data. IEEE Trans Industr Inf 14(7):3205–3215

Al-Jarrah OY, Yoo PD, Muhaidat S, Karagiannidis GK, Taha K (2015) Efficient machine learning for big data: a review. Big Data Res 87–93

Associates DB (2011, Nov) Big data impacts data management: agreement: the 5 Vs of big data. Retrieved from http://davebeulke.com/big-data-impacts-data-management-the-five-vs-of-big-data/

BMC blog (2018) https://www.bmc.com/blogs/machine-learning-data-science-artificial-intelligence-deep-learning-and-statistics/

Data Processing & Data Processing Methods (2018) https://planningtank.com/computer-applications/data-processing-data-processing-methods

Data processing: meaning, definition, steps, types and methods (2018) https://planningtank.com/computer-applications/data-processing

Dean J, Ghemawat S (2004) MapReduce: simplified data processing on large clusters. In: Proceedings of 6th symposium on operating systems design and implementation, pp 137–149

Deng J, Dong W et al (2009) Imagenet: a large-scale hierarchical image database. In: Paper presented at IEEE conference on computer vision and pattern recognition, Miami, FL, USA, pp 20–25

Explain three methods of data processing (2018) https://www.kenyaplex.com/questions/16977-explain-three-methods-of-data-processing.aspx

Five data mining techniques that help create business value (2018) https://datafloq.com/read/data-mining-techniques-create-business-value/121

Group ES (2012, August) The 6 vs: the BI/analytics game changes so microsoft changes excel. Retrieved from http://www.esg-global.com/blogs/the-6-vs-the-bianalytics-game-changes-so-microsoft-changes-excel/

Jan B, Farman H et al (2017) Deep learning in big data analytics: a comparative study. Comput Electr Eng. https://doi.org/10.1016/j.compeleceng.2017.12.009

Khan M, Han K, Karthik S (2018a) Designing smart control systems based on internet of things and big data analytics. Wirel Pers Commun 1683–1697

Khan M, Iqbal J, Talha M, Arshad M, Diyan M, Han K (2018b) Big data processing using internet of software defined things in smart cities. Int J Parall Program 1–14

Khumoyun A, Cui Y et al (2016) Spark based distributed deep learning framework for big data applications. Paper presented at international conference on information science and communications Technologies, Tashkent, Uzbekistan, pp 2–4

Laney D (2001) 3D data management: controlling data volume, velocity, and variety. Application Delivery Strategies, META Group

Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic M (2015) Deep learning applications and challenges in big data analytics. J Big Data 1–21

Qiu JW, Ding G, Xu Y, Feng S (2016) A survey of machine learning for big data processing. EURASIP J Adv Signal Process 1–16

Shvachko K, Kuang H, Radia S, Chansler R (2010) The Hadoop distributed file system. In: Proceedings of the 2010 IEEE 26th symposium on mass storage systems and technologies 1–10

Silva BN, Khan M, Han K (2017) Big data analytics embedded smart city architecture for performance enhancement through real-time data processing and decision-making. Wirel Commun Mob Comput 1–12

Sukumar SR (2014) Machine learning in the big data era: are we there yet? In: Proceedings of the 20th ACM SIGKDD conference on knowledge discovery and data mining: workshop on data science for social good

Towards Data Science (2018) https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d

Vorhies W (2014, October) How many V's in big data? The characteristics that define big data. Data Science Central. Retrieved from http://www.datasciencecentral.com/profiles/blogs/how-many-v-s-in-big-data-the-characteristics-that-define-big-data

Chapter 2
Big Data Analytics

Bhagya Nathali Silva, Muhammad Diyan and Kijun Han

Abstract During the last few decades, with the emergence of smart things and technological advancements of embedded devices, Big Data (BD) and Big Data Analytics (BDA) have been extensively popularized in both industrial and academic domains. The initial portion of the chapter aims to deliver a generic insight toward BD and BDA. In later sections, details that are more specific to BD and BDA are discussed. In fact, BD notion is characterized by its distinctive features such as large amounts of data, high-speed data generation, and wide variety among data type and sources. Consideration on these characteristics assists in determining potential data processing techniques. Hence, this chapter further elaborates on key BD analytical scenarios. Moreover, application of BD, BD analytical tools, and data types of BD are described, in order to enlighten the readers about this broad subject domain. Finally, the chapter concludes by identifying potential opportunities as well as challenges faced by BD and BDA.

List of Abbreviations

IoT	Internet of things
BD	Big Data
BDA	Big Data analytics
ML	Machine learning
DL	Deep learning
MR	MapReduce
TB	Terabyte
PB	Petabyte
EB	Exabyte
ZB	Zettabyte
YB	Yottabyte
RDBMS	Relational database management system
XML	Extensible Markup Language
WSN	Wireless sensor networks

M. Khan et al., *Deep Learning: Convergence to Big Data Analytics*, SpringerBriefs in Computer Science, https://doi.org/10.1007/978-981-13-3459-7_2

CPS	Cyber physical systems
MM	Multimedia
AI	Artificial intelligence
DM	Data mining
OLAP	Online analytical processing
BPM	Business performance management
NLP	Natural language processing
NER	Named-entity recognition
DBMS	Database management systems
URL	Uniform resource locator
NoSQL	Not only SQL
HDFS	Hadoop Distributed File System

2.1 Overview

The emergence of smart devices and connected networks has pioneered Internet of things (IoT) notion, boosting the data generation speed during past few decades (Khan et al. 2016, 2017; Silva et al. 2017b). As reported by International Data Corporation (IDC) in 2011, the annual data volume has increased approximately nine times within five years reaching up to 1.8ZB and further they estimated the data volume growth to be double in every other two years until 2020 (Gantz and Reinsel 2011).

Figure 2.1 concisely illustrates the evolution data growth during past few decades. Big Data (BD) notion was coined as a result of massive data deluge from a wide range of operational domains, i.e., Internet, sensor networks, managerial systems, finance systems, and user-generated data. Since initial divulgement, BD has been in the spotlight alluring both technical experts and public in general and hence been defined by many domain experts considering diversified aspects and perspectives (Silva et al. 2018b). In the beginning, BD was used as a term to define prodigious datasets. However, BD is solely not about the size or the amount of data. Doug Laney defined BD as a 3V model considering distinctive BD characteristics, namely volume, velocity, and variety (Khan et al. 2018; Laney 2001; Silva et al. 2017a). Sheer data size is indicated by volume, whereas velocity term is used to characterize expeditious data creation, and variety defines the diversity among different data sources and data types. Another widely accepted definition by IBM amended 3V model into a 4V model by introducing veracity as the fourth V of any BD model (Gandomi and Haider 2015). Veracity indicates the uncertainty of voluminous data (Khan et al. 2018).

Rapid growth of digital data has driven the research community to data-driven experiments, influencing all aspects of the social dynamics in the society (Silva et al. 2018a). Nevertheless, gathering enormous data amounts will be futile without proper knowledge discovery mechanisms. With rapid data growth, voluminous data analysis for knowledge discovery has become tedious and challenging for data engineering

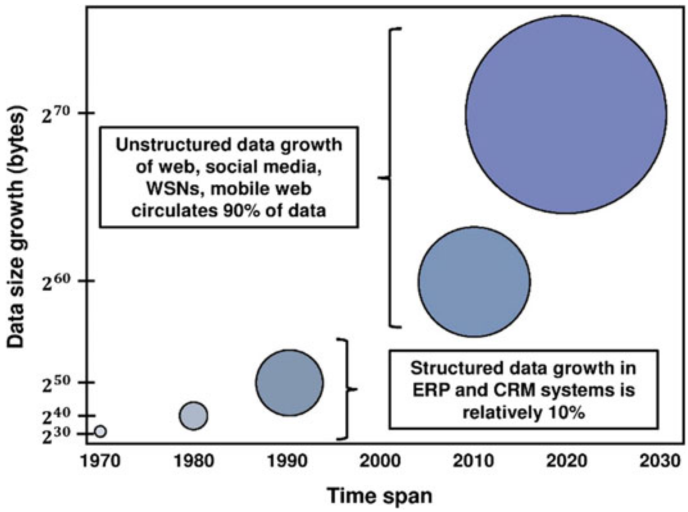

Fig. 2.1 Data growth evolution from 1970 to 2020 (predicted)

experts. Extreme variation among data sources and data types of BD has further exacerbated colossal data management tasks. Hence, BD notion was extended toward Big Data Analytics (BDA) that aims to discover valuable knowledge from large datasets enforcing intelligent decision-making. Therefore, organizations aim to possess efficient data processing mechanisms that transform high-velocity heterogeneous data into valuable information. As reported by Labrinidis and Jagadish (Labrinidis and Jagadish 2012), valuable data extraction from BD can be classified into five stages as shown in Fig. 2.2. These five stages can be bundled as data management and data analytics (Gandomi and Haider 2015). Data management involves data acquisition, storing, preparing, and retrieving for analytics. Data analytics refer to various methods, techniques, and algorithms that extract intelligence from BD. Thus, analytics can be identified as a subprocess of entire BD processing task.

Nevertheless, conventional data analyzing techniques and algorithms fail to process high-velocity data, thus creating a compelling demand to associate other related technologies capable of incorporating advance knowledge discovery mechanisms. Owing to the collaboration among multiple domains and enhanced computation facilities, knowledge discovery process has been broadened and strengthened. For example, domain experts of machine learning (ML) and data engineering have identified the potential of converging ML and BD to improve BDA performance. Contrasting to shallow learning techniques in conventional analytics, ML-converged BDA employs deep learning (DL) techniques to discover intelligence from BD.

As stated afore, BDA aims to explore hidden knowledge encapsulated in undiscovered patterns and correlations (Hu et al. 2014). Considering the demand of time, BDA is categorized into real-time analytics (streaming/online) and archived analytics (batch/off-line). Real-time analytics presumes data value is correlated with data

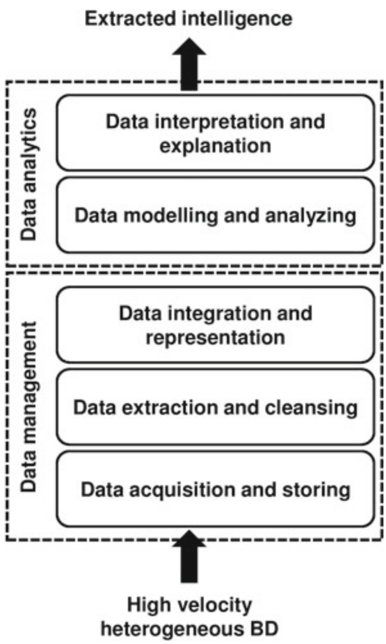

Fig. 2.2 Stages of big data processing to extract valuable knowledge from big data

freshness (Tatbul 2010). In streaming analytics, data arrives continuously as a stream at a high speed. Due to memory constraints, particularly small data portions from the stream are stored and examined to determine potential knowledge from the approximated patterns. Streaming analytics are widely used for real-time applications that require higher responsiveness with utmost accuracy. Spark, Storm, and Kafka are few dominant open-source systems, which support streaming analytics. Contrasting to online data analytics, batch processing analyzes data after storing. MapReduce (MR) is the most widely used batch processing method (Silva et al. 2017a). MR divides a large dataset into smaller portions, in order to perform parallel and distributed processing on each portion. Intermediate results obtained by small portion analysis are combined together to determine the end result.

Owing to potential applicability of discovered patterns and knowledge, many organizations and industries have welcomed and embedded BDA to the organizations' operational frameworks. Consequently, BDA problems are involved in fields varying across social administration, world economy, scientific research, etc. (Chen and Zhang 2014). On the one hand, as reported in McKinsey's report by Manyika et al. (2011), BDA has the competence to flourish global economy with its active participation in various fields. On the other hand, the McKinsey's report (Manyika et al. 2011) claims improvements in social administration can be achieved by introducing BDA functionalities like pattern identification and knowledge mapping to public services sector.

The rest of this chapter will provide descriptive information related to BD characteristics, BD analysis problems, data processing methods, opportunities and challenges, BD applications, data types of BD, and BD tools.

2.2 Characteristics of Big Data

Even though BD term initially coined around mid-1990s, the term became popular around year 2011 creating the hype about BD, owing to extensive efforts made by prominent technology organizations including IBM (Gandomi and Haider 2015). However, with wider acceptance, BD definitions have evolved rapidly inducing rather confusions. Despite of these definition variations, Laney's 3V model (Laney 2001) based on fundamental characteristics of BD was publicly accepted by industrial experts and academia. Since then, the three Vs model, namely volume, velocity, and variety, has been considered as core dimensions of BD management.

As the term implies, volume refers to the size of data. In 2012, IBM conducted a survey with 1144 participants to know about BD familiarity and found out a majority, which is over a half of the participants perceive BD as data with a size beyond one terabyte (TB) (1 TB $= 2^{40}$ Bytes). With expeditious data generation, TB era has evolved to petabyte (PB), exabyte (EB), zettabyte (ZB), and yottabyte (YT). As reported in (Beaver et al. 2010), about 16 million photographs are processed within a second at Facebook, which is approximately equal to 1 TB. According to estimated reports, Facebook stores over 260 billion photographs that roughly span over 20 PBs (1 PB $= 2^{50}$ Bytes). It is worthy to note that magnitude definitions of BD are proportional and correlate with corresponding attributes such as data type, time, and source. Hence, datasets presumed to be BD in the past might not be perceived as BD in the present with respect to current storage capacities and data acquisition methods. Similarly, BD in current era might not be able to surpass the magnitude threshold levels of BD in the future. The type of data is another concern when defining BD level thresholds. Even though data sizes can be equal for two datasets including a tabular dataset and a video dataset, the data type determines the number of records per each dataset. Hence, we can understand that defining a threshold level for data volume in order to interpret BD is impractical.

With the advancements in communication technologies, smart devices, and sensor networks, data acquisition process encountered with heterogeneous data types. Extreme heterogeneity of data in a dataset describes the variety of characteristics of BD. Consequent to technological advancements, data can be structured, semi-structured, or unstructured. Data in relational database management systems (RDBMS) and spreadsheets is identified as structured data as it complies with strict standards to support machine readability. However, only a small portion roughly about 10% belongs to structured data category. Data types with lack of structural organization, i.e., image, audio, video, and text, are categorized as unstructured data. Semi-structured data loosely spans across structured and unstructured types. Extensible Markup Language (XML) is a common semi-structured data form popular with

Web data exchange. Semi-structured data does not adhere to rigorous standards. For example, users define data tags of XML documents, in order to support machine readability. In modern world, many organizations stockpile unstructured data from sensor networks and social media. Even though data hoarding is not a relatively new concept, the novelty and beauty comes with new analytical technologies that improve business processes with stockpiled data. For example, retailers could determine customers' buying pattern and store traffic details exploiting face recognition techniques. This valuable information aids organizations to make lucrative decisions such as staff management, good placement, and appealing promotions.

Velocity implies the data generation rate, and corresponding data analyzing speed requires to make appropriate decisions. The extraordinary data generation speed consequent to the emergence of smart devices and sensor networks has drawn the attention toward real-time data analytics and data-driven decision-making. Data analytics would have to analyze thousands of streaming sources as the smart device usage escalates exponentially. Nevertheless, conventional data processing techniques fail to handle enormous data processing tasks with simultaneous data feeding. As a solution, BDA technologies start to act the role of conventional data processing techniques, with considerably very large datasets. It is noteworthy that BDA technologies are far more efficient than the conventional techniques as they discover real-time intelligence from colossal amount of raw data as batches as well as in streaming form.

In addition to these key attributes, few other characteristics of BD, i.e., veracity, variability, and value, were identified by prominent technology organizations. Veracity defined by IBM is the fourth V of 4V model. This implies uncertainty and unreliability factors of BD. Variability dimension was defined by SAS institute to indicate data flow rate variations that change from time to time with peaks and troughs. In addition to flow rate variations, this attribute encloses complexity factor arose with heterogeneous data sources. Oracle introduced value as another characteristic of BD. In general, BD value is low and not directly proportional to the volume. Nevertheless, extensive analysis on enormous data volumes is promising to improve data value.

2.3 Big Data Processing

As discussed previously, BD is less valuable in its original form. Value-added intelligence could be derived through proper data analytics. However, intelligence discovery from bulks of data is a tedious and challenging task and involves many other stages. This section divides the data processing task into three stages, namely data acquisition, preprocessing, and analyzing as shown in Fig. 2.3.

Data acquisition is the initial stage, which gathers raw data from various real-world objects, devices, and sensors. A well-designed data collection process ensures the validity of results subsequent to data processing procedures. It is worthy to note that data collection design should consider the characteristics of data generators as well as succeeding data analytic objectives. Sensors, log files, and Web crawlers are

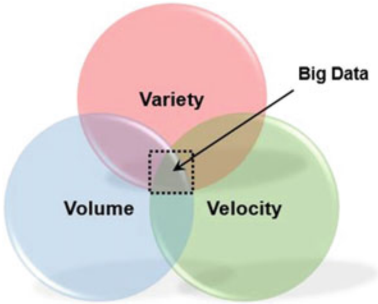

Fig. 2.3 Laney's 3V model based on fundamental characteristics of BD

some common data acquisition techniques (Hu et al. 2014). Sensors are commonly used to acquire physical measurements, e.g., temperature, humidity, and particle concentration in machine-readable digital format. Extended sensor networks can be either wired or wireless. However, during the past few decades, wireless sensor networks (WSN) have gained a significant popularity over wired sensor networks. A system consists a collection of sensors which is known as a cyber physical system (CPS) (Shi et al. 2011). Log files are another widely utilized data acquisition method that records activities in a predefined file format. For instance, Web server log files keep track on every activity performed by Web users. Except from Web servers, stock applications, financial applications, traffic management systems, and many other applications use log files for data acquisition. Web crawler is software that stores Web pages for search engines (Cho and Garcia-Molina 2002). Web crawlers perform data collection in Web-based applications. There are many other data acquisition methods other than afore-stated techniques. Descriptive explanation on different BD data types can be found in (Chen et al. 2014).

Once data is acquired, gathered data undergoes preprocessing, in order to address variabilities in noise, consistency, and redundancy resulted from heterogeneity of sources. The demand for data preprocessing has become significant with the rising cost on data transmission and storage facilities. In order to maintain the quality of stored raw data, preprocessing techniques were introduced to the BD systems. Integration, cleansing, and redundant data elimination are some common preprocessing techniques used in BD systems. Data integration combines data acquired from various sources to facilitate a collective view of data (Lenzerini 2002). Even though conventional data management incorporates data warehouse method and data federation method, "store-and-pull" characteristics of these two techniques hinder their applicability with high-velocity real-time applications. Therefore, utilizing integration techniques on streaming engines (Tatbul 2010) and search engines (Cafarella et al. 2009) would be beneficial. Data cleansing identifies uncertain, incomplete, or ambiguous data and modifies or removes them from the dataset, subsequently improving the quality of data. Considering the complexity of cleansing techniques, embedding them should be done with care to obtain data accuracy, without compromising the computational speed. Redundancy elimination techniques reduce the amount of repeated and superfluous data to improve data transmission, storage utiliza-

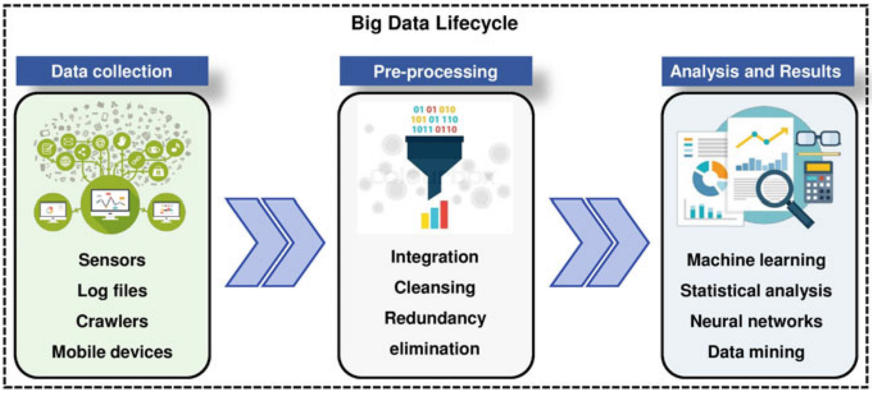

Fig. 2.4 Stages of big data life cycle with corresponding technologies

tion, reliability, and data consistency. Similar with all other approaches, redundancy elimination comes with benefits and trade-offs. Therefore, mindful usage can improve the data quality without straining the system on data compression and decompression procedures. In fact, none of the techniques assures optimal preprocessing over a wide range of large datasets. Hence, collective consideration of data characteristics, problems to be solved, and analytic objectives would be crucial for the selection of proper preprocessing mechanism (Fig. 2.4).

Data analysis is the most imperative stage of deriving intelligence from BD. The utmost goal of data analysis is to discover hidden knowledge that supports decision-making. In accordance with Hu et al., data analytics can be divided into six major research areas, i.e., structured data analytics, text analytics, Web analytics, network analytics, multimedia (MM) analytics, and mobile data analytics (Hu et al. 2014). Even though the objective of data analysis varies across different domains, some data analyzing approaches, i.e., data visualization, machine learning, neural networks, statistical analysis, pattern recognition, signal processing, and data mining, are beneficial and convenient to use in every domain (Khan et al. 2018). Data visualization methods deliver valuable information in the form of graphical illustrations. BD visualization has been studied by many interest groups, owing to the benefits in software designing and algorithm development. ML is a part of artificial intelligence (AI) that designs algorithms to learn computational behaviors with experimental data. Not only that ML enables automatic knowledge discovery and intelligent decision-making, the consolidation of ML with deep learning has pioneered deep machine learning (DML) as a novel research frontier. Statistical analysis is another common data processing method that collects, organizes, and interprets data. The statistical analysis approach identifies general relationships and valuable correlations among various data attributes. Nevertheless, lack of suitability of general statistics in BD analytics has paved the way for parallel statistical analysis and large-scale statistical algorithms. Data mining (DM) is highly favored in BD processing to extract hidden knowledge from very large datasets. In other words, DM can be expressed as a consolidation of ML and statistics. BD mining is more challenging than conventional DM and required to be dealt with gigantic datasets. According to Wu et al. (2008), Apriori, K-means, Naïve Bayes, Cart, etc., were identified as promising DM algorithms for BD.

With the rising of BD era, experts were keen on extracting valuable data rapidly, using extensively large datasets. For that purpose, many organizations use hashing, indexing, bloom filtering, and parallel computing techniques. Hashing transforms data into index values or fixed-length numerical values to increase the speed of reading, writing, and querying, Indexing is used to reduce the cost of reading and writing by boosting the speed of insert, delete, update, and query processes. The beauty of indexing is that it can be applied on all forms of data including structured, semi-structures, and unstructured data. Bloom filter is another form of BD processing technique, which consists of a collection of hash functions along with data to facilitate lossy data compression. With increasing demands in data processing, many computational resources act together simultaneously in parallel computing. Contrasting to conventional serial computing, parallel computing segregates a computational task into several smaller tasks and assigns them to different computing resources to achieve co-processing.

2.4 Data Analysis Problems in Big Data

As stated in Sect. 1.3, BD analytical problems can be broadly categorized into six major areas, namely structured data analytics, text analytics, Web analytics, multimedia analytics, network analytics, and mobile data analytics.

Business organizations and scientific communities generate a colossal amount of structured data. Structured data analytics highly rely upon RDBMS, online analytical processing (OLAP), business performance management (BPM), and data warehousing. Deep learning has become an active player in structure BD analysis. DL-incorporated ML algorithms play a significant role in learning multilevel data representation models (Hinton 2007). Furthermore, mathematical model-based ML is occupied in sophisticated structured application domains such as energy management and fault detection (Baah et al. 2006; Moeng and Melhem 2010).

Text is a common form of data stored in documents, Web pages, emails, and social media. Considering the widespread nature, text analytics is considered to be vital than conventional structured data analytics. Text mining is another term that refers to text analytics, which discovers knowledge from unstructured textual data. Text analytical problems involve ML, statistics, DM, computation linguistics, and natural language processing (NLP) techniques. NLP techniques empower computers to comprehend text, analyze textual data, and generate textual information. Automatic knowledge extraction involves information extraction based on a specific topic or a theme. Named-entity recognition (NER) is widely utilized in such scenarios to determine data affiliated with predefined categories. Extractive summarization is another solution approach for text analytical problems, which summarizes a report to few key sentences. Graph-based text mining is also used to categorize textual data according to a theme or a topic.

Consequent to the eruptive growth in number of Web pages, Web analytics came to the spotlight during last two decades. Knowledge discovery and intelligence extraction from Web resources is considered to be the utmost goal of Web analytics. Web analytics involve data retrieval, extraction, and data evaluation. Web analytics is further categorized into content mining, usage mining, and structure mining (Pal et al. 2002). Web content mining utilizes either database approach or information retrieval approach (Hu et al. 2014). Mining of secondary data generated by Web sessions belongs to Web usage mining. Usage mining considers user details, user registration, user queries, cookies, logs from server, browser, and proxy. Discovering graph links within a Web site or among different Web sites defines Web structure mining.

With the phenomenal popularity of social media, MM data grew rapidly along with ubiquitous accessibility. Consequently, multimedia analytics was emerged as a novel research area that aims to extract new knowledge from any form of multimedia data, i.e., audio, video, and image. MM analytics is challenging and tedious due to variability in multimedia data types. Nevertheless, MM data contains more information than structured data and textual data. Hence, experts in both academia and industry have proposed and studied many interest areas, i.e., MM annotation, retrieval, indexing, summarizing, that address key challenges of MM analytics. MM annotation assigns labels to MM files outlining the file contents. Indexing and retrieval techniques assist people to discover intending MM file easily and swiftly. MM summarization derives the most notable audio content or video content that best describes the original file.

Explosive growth in connected networks and social networks has extended conventional bibliometric network analytics to sociology network analytics (Watts 2004). Social networks consist of content data (text, image, video, etc.) and graph-like linkage data. In fact, content data and linkage data of social media are exceptionally rich from hidden knowledge. However, extreme complexity of these lucrative data has become a challenge for social network analytics. Social network analytics comprises of two major research areas, namely content data analytics and linkage data analytics. More elaborated explanations on content analytics and linkage analytics can be found in (Aggarwal and Wang 2011). As stated, social media consists of heterogeneous data types; hence, every BD analytic approach can be incorporated in social network analytics. Nevertheless, these approaches should be tailored to meet time constraints, noisiness of data, and dynamicity attributes of social networks.

Mobile analytics emerged with tremendous growth in mobile data, subsequent to advances in mobile computing. However, mobile data analytics encounter with challenges result from mobile data characteristics such as noisiness, data redundancy, location awareness, and activity awareness. In order to ensure user satisfaction, mobile analytics should facilitate intelligent decision-making on real-time basis. Owing to knowledge discovery served with mobile analytics, building of complex mobile systems that use empirical data has become more reliable and easy.

2.5 Applications of Big Data

BD applications assist small- and large-scale businesses to make intelligent decisions with the aid of voluminous data analysis (Hilbert 2016; Sagiroglu and Sinanc 2013). Internet clicks, Web server logs, activity reports, content of social media, mobile phone history records, email texts of customers, and sensor data are the main sources of data. Different interests of organizations enforce BD applications to disclose hidden information by exploring large volumes of data records. This section will cover BD applications in different domains (Clegg 2009).

2.5.1 Healthcare

Data generated in healthcare systems is not insignificant. Due to inadequate capacity and shortcomings in standardization, the healthcare industry covers BD. The usage of BDA in healthcare has revolutionized to make personalized medication easy and smart. For particular diseases, researchers are in race to provide suitable treatments in specific time intervals, analyze data pattern to recognize drug side effects and future treatments, and lower the treatment costs. Sophisticated views on health, i.e., eHealth, mHealth, and wearable devices, generate prodigious amount of data including images, records, sensor data, and patient data. Another potential application of BD in healthcare is to predict specific regions affected by some diseases by mapping health and geographical data. Consequent to accurate and timely predictions, doctors can act proactively to plan, diagnose, and provide vaccines and serums.

2.5.2 Manufacturing

A key potential benefit of applying BDA in production and manufacturing domains is to gain transparency and zero downtime. For this purpose, large amount of data is required together with latest analytical tools, in order to process data on real-time basis and to extract valuable knowledge from heaps of raw data. Following are the main areas in manufacturing and production domain, which utilize BD application for performance enhancements.

- Productivity and tolerance tracking
- Planning and Management in supply chain
- Energy efficiency
- Processes of testing and emulation for new products
- Customization for massive manufacturing

2.5.3 *Government*

The performance of government operations could be improved by the adaptation of BDA concepts. A particular set of data is shared among multiple applications as well as multiple governing departments to extract knowledge. Innovative applications play an important role in e-governance in multiple domains. Following are some of the areas of governance where BD and BDA play their roles for performance enhancements.

- Cyber security and intelligence
- Tax compliance
- Crime identification and its prevention measure
- Traffic management
- Weather
- Research and development

2.5.4 *Internet of Things*

IoT is another wide application domain of BD, which on the other hand is a source of BD creation with its enormous number of connected sensors, things, and devices. IoT gathers sensor and actuator data and then uses them in a variety of contexts.

2.6 Data Types of Big Data

A variety of data types exists in BDA, namely network data, linked data, event data, time series data, natural language data, real-time media data, structured data, and unstructured data. With the unceasing data growth, distinguishing valuable data has become a crucial demand in modern scientific era. Nevertheless, determining valuable data from ambiguous, noisy, and misinterpreted is still being a challenge to data scientists all over the world. This section provides concise descriptions about afore-stated data types.

1. Structured data: Numerical data stored in rows and columns, where every data element is defined, is known as structured data. About 10% of total data volume belongs to this type of data, which is accessible through database management systems (DBMS). Governmental departments, real estate and enterprises, organizations, and transactions generate structured data in each of their operational processes.
2. Unstructured data: Anything apart from tabular data with defined data elements belongs to unstructured data category. Data in the form of image, audio, video, and text is considered as unstructured data. This type of data contributes up to 90%

of the total data volume. In last few decades, popularity of social media has added more unstructured data, which cannot be stored and processed using traditional data storing and processing approaches. Such data can be stored in appropriate databases. NoSQL system is one of the suitable storage for unstructured data. CouchDB and MongoDB are widely used by many organizations as NoSQL repositories.

3. Geographic data: Data generated from geographic information systems (GISs), i.e., address, workplaces, buildings, transportation route, and road systems, belongs to geographic data. These data elements are easy to gather via deployed sensors and GIS. Consequent to data gathering, altering, processing, and analyzing tasks take place in order to derive knowledge from raw data. Moreover, geostatistics are occupied to monitor environmental conditions.
4. Real-time media: Streaming live media data or stored media data results in real-time media. In fact, storing and processing real-time media has been a challenge due to its streaming nature. There are number of sources that generate audio, video, and image data. To name a few, flicker, YouTube, and Vimeo are among the key real-time media generating sources. Videoconferencing is the other source of real-time media data, which facilitates seamless full-duplex communication.
5. Natural language data: People generate data in the form of verbal conversations. The level of editorial quality and level of abstraction are different of such type of data. Speech capturing devices, Internet of things, mobile phones, and fixed-line phones are the main sources of natural language data.
6. Time series: Measurements or observations on a particular indexed in time order are considered as time series. In general, time series data is observed at equal time intervals. Time series data is analyzed to extract hidden knowledge from gathered time data. Signal processing, statistics, earthquake prediction systems, pattern recognition, and many more other areas utilize time series data type.
7. Event data: Event data is generated with the correlation between external data and time series data. The utmost goal of event data type is to distinguish valuable events among innumerable amount of events taking place. The three components of an event data are action, time interval, and state. Combination of all three components creates an event data. Event data represents nested, renormalized, and schema-less characteristics.
8. Network data: Large networks such as twitter, YouTube, and Facebook generate network data. Other information networks, biological networks, and technological networks are also act as sources of networks data. Network data can be represented as one-to-one or one-to-many relationships among network nodes. A node can be a user, data item, internet device, neural cell, etc. Maintaining connection between nodes and network structures is considered to be the main challenge for network data.
9. Linked data: Linked data type includes URLs in Web technology. Computers, embedded devices, and other smart devices are then able to share and inquire semantic information. Data can be read and shared using linked data types.

2.7 Big Data Tools

Continuous development of business operations highly relies on thorough investigation of BD. Data analytics plays an important role in intelligent decision-making as well as development and well-being of the organization. However, processing sheer amounts of data with the aid of conventional data processing tools is not feasible and not efficient. Hence, various BD tools were introduced in the recent past. BD tools assist organizations as well as data scientists to derive knowledge-driven decisions efficiently and cost effectively. A variety of BDA tools are being used by the practitioners to facilitate data storage, data management, data cleansing, data mining, data visualizing, and data validation. This section briefly outlines widely used BDA tools.

1. NoSQL

In general, SQL is widely used to handle and query structured data. However, tremendous growth of unstructured data has pioneered the emergence of unstructured data analytical tools. Subsequently, not only SQL (NoSQL) has been developed to handle unstructured data types effectively. NoSQL databases do not particularly adhere to a schema when storing unstructured data. Hence, the column values of the table vary with respect to each data record (row). Due to this schema-less nature, NoSQL storages compromise consistency over speed, fault tolerance, and availability. Even though NoSQL has gained immense popularity during last few decades, challenges arise from low-level query languages and lack of standardized interfaces is yet to be addressed.

2. Cassandra

Apache Cassandra is a NoSQL, open-source, distributed database that is capable of managing very large datasets across multiple servers. The beauty of Cassandra is that it comes with no single point of failure assuring high availability under any circumstances. Hence, Cassandra is being highly favored by expert groups when scalability and availability features are crucial without compromising performance. Moreover, Cassandra facilitates data replicating across multiple clouds or data centers to ensure lower latency and fault tolerance.

3. Hadoop

Hadoop is a framework that consists of a collection of software libraries that incorporate various programming models to facilitate distributed processing of large datasets. Scalability is another benefit comes with Hadoop. Distributed storage facility offered with Hadoop is separately known as Hadoop Distributed File System (HDFS). The Hadoop does not rely on hardware to facilitate high availability. Instead, it incorporates software libraries to detect, identify, and handle points of failures at the application layer. The Hadoop framework consists of Hadoop Common (common libraries), HDFS (storage), Hadoop YARN (schedule computing resources), and Hadoop MapReduce (process large datasets).

4. Strom

Storm is a real-time open-source distributed computation system. Storm manages large amount of streaming data, which is similar to batch processing of Hadoop. The beauty of Storm is that it can be handled with any programming language. In order to ensure proper real-time analytics experience, Storm integrates the existing queueing mechanisms as well as database technologies. Streaming processing of Storm requires complex and arbitrary partitioning at each computation stage. Real-time analytics, unceasing computation, and online ML are some of the key services offered by Storm.

5. Spark

Spark is a general-purpose open-source distributed cluster-computing framework. Spark guarantees higher performance in both stream data processing and batch processing. Spark integrates SQL, GraphX, MLib, and Spark streaming components. Spark can access data from various data sources and run on various platforms such as Hadoop, Mesos, and Yarn. In comparison with Storm, Spark is considered cost effective, since same code set can be used for both real-time processing and batch processing. However, Storm has gained superiority in terms of latency with fewer restrictions.

6. Hive

Hive is a cross-platform data warehouse software operates on Hadoop. Hive enables data analysis and querying of data stored in multiple storages and file systems that are integrated to Hadoop. In order to query over distributed file systems, Hive provides SQL abstraction via HiveQL, which is a SQL-like querying language. Hence, applications do not have to implement queries using low-level Java APIs. HiveQL transparently converts queries into Spark, Apache Tez, and MapReduce. Moreover, Hive offers indexes to improve the query execution speed.

7. OpenRefine

OpenRefine is an open-source stand-alone application that was previously known as Google Refine. OpenRefine is widely used in present data-driven world to cleanse noisy raw data and to transform data from one form to another. OpenRefine is similar to typical RDBMS tables in a way as it stores rows of data under columns. However, it deviates from the conventional scenario, as no formulas are stored in cells under columns. The formulas are used in OpenRefine to transform data, and they are transformed once only.

2.8 Opportunities and Challenges

As in every other field, BD owns its benefits, advantages, and opportunities, which are followed by challenges. With prominent opportunities in one hand, on the hand BD analytics face vital challenges. This section briefly discusses the potential opportunities and identified challenges.

National Science Foundation (NSF) and National Institute of Health (NIH) of USA have recently confirmed that employing BD analytics in their decision-making process has proven the potential benefits for future endeavors (Chen and Zhang 2014). In fact, BD era has changed everyone's lifestyle with its impacts on social and economic development. Rise of BD analytics transformed people's behaviorism from reactive to proactive, subsequently enforcing precautions for possible events that might take place in future. Owing to the benefits gained through intelligent knowledge discovery, BD will become the playmaker for many organizations to attract skillful employees, while assisting in critical decision-making to compete with other competitors. With the emergence of BD analytics, Hadoop technologies have attained a remarkable popularity and success. Nevertheless, in future, these technologies might not be sufficient to manipulate high-velocity BD. Hence, focusing on sophisticated processing techniques and storage technologies such as distributed databases would create a significant breakthrough. Moreover, coexisting operation of BD analytics with other technologies such as IoT, mobile computing, and cloud computing innovates technological advancements to assist efficient data acquisition, data storage, data analysis, and data protection. Novel data representation methods are another realm of opportunities that goes together with the maturation of BD analytics, since illustration of analytical results is one of the most factors in intelligent decision-making. Extending BD research opportunities is likely to be heightened in future, as BD is considered as an extension of human brain rather than a brain substitution.

BD analytics face different challenges throughout the BD life cycle. To be specific, challenges in data collection, data storage, data processing, data analyzing, results representation may suppress the full capacity of BD analytics. Extreme complexity of BD has diminished the applicability and continuous advancements of analytical algorithms. Sluggish evolution of BD analytics id always accompanied by BD characteristics is denoted by 3V model. Large volume datasets generate countless number of output classes and free parameters that adversely affect the processing time and complexity. In order to manage these challenges, experts have extended BD analytics to make fusion with parallel computing that incorporates clusters of CPUs and GPUs. Complexity of BD analytics is highly influenced by the heterogeneity of data types. Hence, BD analytical techniques should strictly focus on mechanisms that alleviate challenges arose with extreme data variety. Integrating heterogeneous data types for machine learning processes is another key challenge encounter with variety characteristic of BD. BD analytics is further challenged by rapid data generation rates. High-velocity data requires real-time processing to serve decision-making. However, minimal works have been performed on online learning process with BD, thus requiring more developments. Experts foresee mini-batch processing supported by parallelism as promising technique to manage high-velocity BD analytics. Extreme non-stationary nature associated with high velocity is another hindrance factor for real-time BD analytics.

After the thorough literature review, we identified afore-stated opportunities and challenges attached with BD. Consequent to the unceasing technical growth, many of the experts asserted to be optimistic about the potential benefits of BD analytics. Accordingly, they claim that technological advancements would be able to overcome the obstacles arose with BD volume, velocity, and variety.

References

Aggarwal CC, Wang H (2011) Text mining in social networks. In: Social network data analytics. Springer, pp 353–378

Baah GK, Gray A, Harrold MJ (2006) On-line anomaly detection of deployed software: a statistical machine learning approach. In: Proceedings of the 3rd international workshop on Software quality assurance, 2006. ACM, pp 70–77

Beaver D, Kumar S, Li HC, Sobel J, Vajgel P (2010) Finding a needle in haystack: Facebook's photo storage. In: OSDI, 2010, pp 1–8

Cafarella MJ, Halevy A, Khoussainova N (2009) Data integration for the relational web. Proceed VLDB Endowment 2:1090–1101

Chen CP, Zhang C-Y (2014) Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. Inf Sci 275:314–347

Chen M, Mao S, Liu Y (2014) Big data: a survey. Mob Netw Appl 19:171–209

Cho J, Garcia-Molina H (2002) Parallel crawlers. In: Proceedings of the 11th international conference on World Wide Web, 2002. ACM, pp 124–135

Clegg B (2009) Before the Big Bang: the prehistory of our universe. St. Martin's Press

Gandomi A, Haider M (2015) Beyond the hype: Big data concepts, methods, and analytics. Int J Inf Manage 35:137–144

Gantz J, Reinsel D (2011) Extracting value from chaos. IDC iView 1142:1–12

Hilbert M (2016) Big data for development: a review of promises and challenges. Dev Policy Rev 34:135–174

Hinton GE (2007) Learning multiple layers of representation. Trends Cogn Sci 11:428–434

Hu H, Wen Y, Chua T-S, Li X (2014) Toward scalable systems for big data analytics: a technology tutorial. IEEE Access 2:652–687

Khan M, Silva BN, Han K (2016) Internet of things based energy aware smart home control system. IEEE Access 4:7556–7566

Khan M, Silva BN, Jung C, Han K (2017) A context-aware smart home control system based on ZigBee sensor network. KSII Trans Internet Inf Syst 11:1057–1069

Khan M, Silva BN, Han K (2018) Efficiently processing big data in real-time employing deep learning algorithms. In: Deep learning innovations and their convergence with big data. IGI Global, pp 61–78

Labrinidis A, Jagadish HV (2012) Challenges and opportunities with big data. Proceed VLDB Endowment 5:2032–2033

Laney D (2001) 3D data management: controlling data volume, velocity and variety. META Group Res Note 6:1

Lenzerini M (2002) Data integration: a theoretical perspective. In: Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems, 2002. ACM, pp 233–246

Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH (2011) Big data: the next frontier for innovation, competition, and productivity

Moeng M, Melhem R (2010) Applying statistical machine learning to multicore voltage and frequency scaling. In: Proceedings of the 7th ACM international conference on computing frontiers, 2010. ACM, pp 277–286

Pal SK, Talwar V, Mitra P (2002) Web mining in soft computing framework: relevance, state of the art and future directions. IEEE Trans Neural Netw 13:1163–1177
Sagiroglu S, Sinanc D (2013) Big data: a review. In: 2013 international conference on collaboration technologies and systems (CTS). IEEE, pp 42–47
Shi J, Wan J, Yan H, Suo H (2011) A survey of cyber-physical systems. In: 2011 international conference on wireless communications and signal processing (WCSP). IEEE, pp 1–6
Silva BN, Khan M, Han K (2017a) Integration of Big Data analytics embedded smart city architecture with RESTful web of things for efficient service provision and energy management. Fut Gener Comput Syst
Silva BN, Khan M, Han K (2017b) Internet of things: a comprehensive review of enabling technologies, architecture, and challenges. IETE Tech Rev 1–16
Silva BN, Khan M, Han K (2018a) Towards sustainable smart cities: a review of trends, architectures, components, and open challenges in smart cities. Sustain Cities Soc 38:697–713
Silva BN et al (2018b) Urban planning and smart city decision management empowered by real-time data processing using big data analytics. Sensors 18. https://doi.org/10.3390/s18092994
Tatbul N (2010) Streaming data integration: challenges and opportunities
Watts DJ (2004) Six degrees: the science of a connected age. WW Norton & Company
Wu X et al (2008) Top 10 algorithms in data mining. Knowl Inf Syst 14:1–37

Chapter 3
Deep Learning Methods and Applications

Jamil Ahmad, Haleem Farman and Zahoor Jan

Abstract This chapter introduces the various methods existing beneath the umbrella of deep learning paradigm, their intricate details, and their applications in various fields. Deep learning has substantially improved the predictive capacity of computing devices, due to the availability of big data, with the help of superior learning algorithms. It has made it possible as well as practical to integrate machine learning with sophisticated applications including image recognition, object detection, self-driving cars, drug discovery, and disease detection. The superior and reliable performance of deep learning methods has attracted the attention of researchers working in every field of science to utilize their strengths in order to solve problems. In addition to that, the knowledge reuse in deep learning is an interesting aspect of this technology which will also be discussed.

List of Acronyms

CNN	Convolutional neural network
DNN	Deep neural networks
RNN	Recurrent neural network
LSTM	Long short-term memory
IR	Information retrieval
BoW	Bag-of-words
CBIR	Content-based image retrieval
NLP	Natural language processing
ML	Machine learning
MTL	Multitask learning

M. Khan et al., *Deep Learning: Convergence to Big Data Analytics*, SpringerBriefs in Computer Science, https://doi.org/10.1007/978-981-13-3459-7_3

3.1 Background

Scientists have been working on deep neural networks since 1979; however, it emerged as a new machine learning research area in 2006 when it was used to reduce the dimensionality of data in an unsupervised manner (Hinton and Salakhutdinov 2006). In 2012, it grabbed the attention of researchers when it won the ImageNet competition beating the runner-up with a huge margin. Deep convolutional neural network (CNN) was employed to automatically extract visual features from a large number of images and then perform image classification (Krizhevsky et al. 2012). Since then, active research is going on in deep learning to solve complex problems in almost all areas of research. Particularly, state-of-the-art performance was achieved in image recognition, speech recognition, semantic image segmentation, natural language processing, and many other tasks.

Deep learning is a set of algorithms in machine learning which attempts to learn important features from raw data automatically. Like traditional neural networks, the deep neural networks (DNN) also consist of artificial neurons arranged in the form of input, hidden, and output layers (Hinton et al. 2012). However, unlike the traditional neural networks, the number of hidden layers in deep networks is usually more than one. The hierarchical nature of deep neural networks allows them to learn features at multiple levels where each level corresponds to a particular level of abstraction. Basic features are learnt at the initial layers which are then aggregated at the deeper layers to construct higher level concepts. It can be applied directly to a variety of data including images, audio, and text (LeCun et al. 2015).

The most important reasons for the popularity of deep learning are the highly improved parallel processing abilities of hardware, especially the general-purpose graphical processing units or GPUs, the substantially increased size of data available for training, and the recent advances in machine learning algorithms. These advances enable deep learning methods to effectively utilize complex, compositional nonlinear functions, to automatically learn distributed and hierarchical features, by effectively utilizing both labeled and unlabeled data.

3.2 Categorization of Deep Learning Networks

Like machine learning, deep learning also encompasses a broad array to approach for learning from data. The hierarchical architectures of deep neural networks can be used in a variety of ways to solve different problems. In general, deep learning can be categorized into three different groups.

3.3 Deep Networks for Supervised Learning

These networks provide discriminative power for pattern classification and regression. Based on the labels provided with the data, the network attempts to learn the difference among data items belonging to various classes. In both classification and regression tasks, the network learns to map input to the expected output (label). Common architectures for supervised learning include DNN, CNN, and recurrent neural network (RNN).

In DNN, each layer consists of a number of neurons, forming a hierarchy. The output of the previous layer becomes the input of the next layer and so on. Each subsequent layer learns increasingly complex patterns in the input data. Lower layers typically learn low-level features, whereas the deeper layers learn high-level abstractions in the data. DNNs are the simplest in term of structure because they are feed-forward neural networks with many layers as shown in Fig. 3.1.

Another famous architecture for supervised learning is CNN which was primarily introduced for visual data processing like images and videos. However, they have proven to be extremely useful for almost any type of data including visual (Ahmad et al. 2018a), audio (Badshah et al. 2017), and even text (Glorot et al. 2011). CNNs consist of three different types of layers including convolutional, pooling, and fully connected layers. The convolution layers attempt to learn significant features that may appear in the data with the help of filters/kernels whose coefficients are tuned during the training phase. Each filter is separately convolved over the input to obtain a feature map, where the location of features is marked with higher activation values (Ahmad et al. 2017a). Like the DNNs, the lower layers in the CNN learn basic features and as we go deeper in the network, the kernels learn more and more complex features. The pooling layers reduce the dimensionality of feature maps and also introduce some

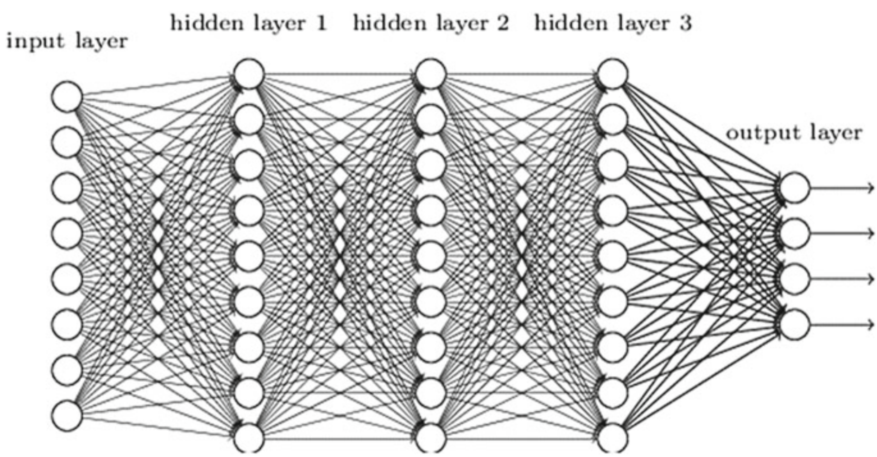

Fig. 3.1 A fully connected feed-forward deep neural network

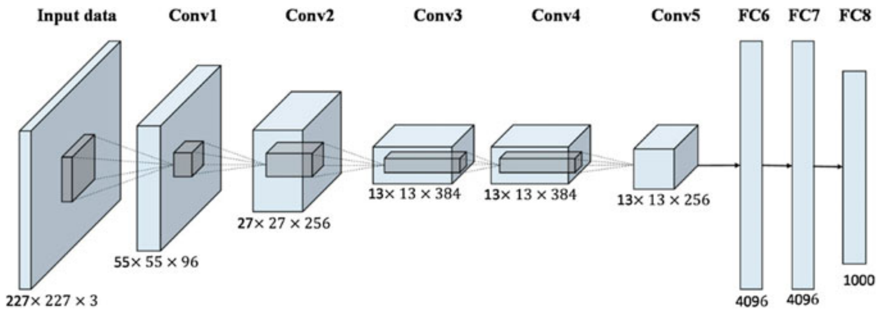

Fig. 3.2 AlexNet convolutional neural network architecture

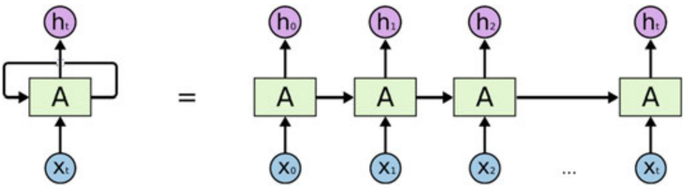

Fig. 3.3 Architecture of a single recurrent unit in RNN

degree of translation invariance in the network. The convolutional and pooling layer forms the features extraction pipeline of the network which detects local features in the input. The fully connected layers then combine the local features to obtain global features (Ahmad et al. 2017b).

CNNs were introduced in 1979; however, these networks got popularity in 2012, when it won the famous ImageNet challenge by a large margin. The network, known as AlexNet consisted of five convolutional layers, three pooling layers, and three fully connected layers as shown in Fig. 3.2.

Both DNN and CNN are very powerful architectures for analyzing non-sequential data; however, they are not suitable for detecting patterns in time series data. For this purpose, a new class of architectures was developed known as recurrent neural network (RNN) (Zaremba et al. 2014). Each unit in the RNN contains recurrent connections in such a way that the network can retain information over a longer period of time. This enables RNNs to recognize patterns in sequential data like speech, videos, and text. A more recent and advanced form of RNN is known as long short-term memory (LSTM) network which improves the pattern recognition ability of RNNs (Ullah et al. 2018). The architectures of RNN and LSTM units are shown in Fig. 3.3.

3.4 Deep Networks for Unsupervised Learning

Unsupervised learning refers to the leaning methods where task-specific supervision information (e.g., target class labels) is not available in the learning process. The most common methods for unsupervised learning include deep autoencoder (Hinton and Salakhutdinov 2006) and deep Boltzmann machines (DBM) (Lu et al. 2017).

Autoencoders and deep bottleneck networks consist of two parts. The first part attempts to compress the input data to a relatively short length representation. The second part is then used to reconstruct the original input from this short representation. During training, the autoencoder attempts to derive such a compact representation which will facilitate the reconstruction of the original data with minimal loss. This way, it learns highly significant features of the training data in an unsupervised manner. The compact representation derived by the autoencoder is often used as a feature vector of the high dimensional input and can serve various purposed including clustering, indexing, and searching, and even dimensionality reduction or feature embedding. A typical autoencoder with the encoder and decoder part is shown in Fig. 3.4.

DBMs are probabilistic generative models which are composed of multiple layers of stochastic, hidden variables. The top two layers consist of undirected, symmetric connections between them, whereas the lower layers receive top-down, directed connections from the layer above them.

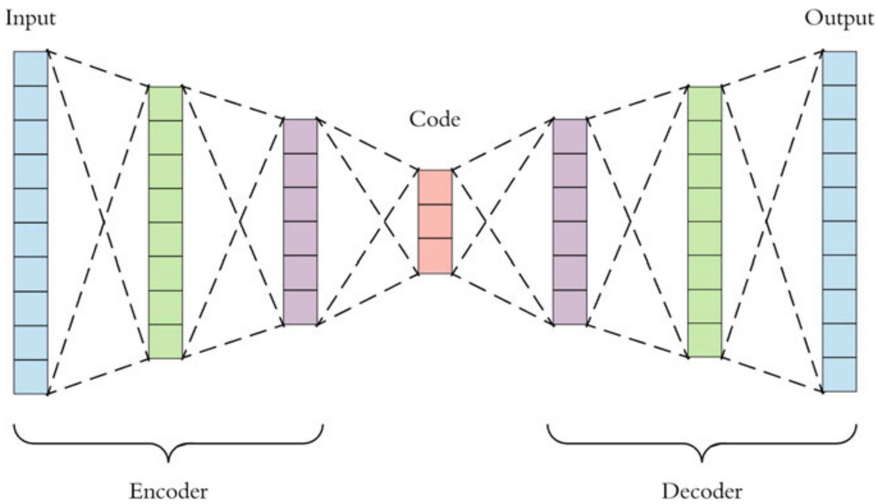

Fig. 3.4 Architecture of a simple autoencoder showing the encoder and decoder parts

3.5 Hybrid Approach

In the hybrid approach, the objective may be discrimination which may often be assisted in a significant way, with the outcomes of generative or unsupervised deep network. It can be accomplished by better optimization and regularization of the deep networks in supervised learning. For instance, a large amount of unlabeled data may be used in unsupervised learning method in order to learn the initial parameters for the later supervised learning task. The approach can also be used when discriminative criteria for supervised learning are used to estimate the parameters in any of the deep generative or unsupervised deep networks using unsupervised learning.

3.6 Transfer Learning Techniques

It is a machine learning method where a model developed for one task is reused as the starting point for a model on another task. It is a popular approach in deep learning where the weight of pretrained models is used to initiate parameters of models for computer vision and natural language processing tasks. Doing so makes it very easy for the training algorithm to optimize the weights by slightly adjusting them according to the new domain. In the field of computer vision, pretrained models obtained from very large datasets like ImageNet are used to initialize classification models, which are then fine-tuned on the target dataset (say radiographs). For instance, Ahmad et al. used a pretrained ImageNet model to classify radiographs by using transfer learning approach (Ahmad et al. 2017b).

Two approaches are commonly employed when using transfer learning on computer vision problems.

Firstly, if a fairly sufficient number of images (e.g., >1,000 images per class) is available, then the weights of the mode can be initialized with the weights of a model trained on a different dataset. During training, most of the layers will be kept unchanged (typically the first convolutional layers) and the parameters of the higher level layers are optimized. The goal of transfer learning in this scenario is to reduce the number of parameters which needs to be optimized, while reusing the lower level layers. The lower level layers are considered as generic feature extractors regardless of the problem domain, and the model has freedom of combining higher level layers together, specific to the problem.

In a second scenario, if the number of available images is small (<1,000), then retraining an existing model will most likely result in overfitting. The number of parameters that needs to be optimized would simply be too large with respect to the number of available images. Regardless, as long as the data is visually similar to the images in a large dataset, then a large pretrained network (trained on that large dataset) can be used as a feature extractor. More specifically, the last N layers of the pretrained network (typically $N = 1$ or $N = 2$) will be removed, and the output of the final deep layer will be used as feature representation of the images. This is again based on the assumption that the first layers in the pretrained network learn problem independent features. These features can then be used with a traditional machine learning or computer vision approach.

3.6.1 Homogenous Transfer Learning

When the source and target domains share similar feature space, it is referred to as homogeneous transfer learning.

3.6.2 Heterogeneous Transfer Learning

In a transfer learning scenario, if the source and target do not belong to the same or similar feature space, then it is known as heterogeneous transfer learning.

3.7 Applications of Deep Learning

Deep learning has been applied to a huge variety of areas like computer vision, speech recognition, entertainment, games, malware detection, fraud detection, education, manufacturing, agriculture, and self-driving cars. The highly capable algorithms along with the availability of big data have opened tremendous opportunities for researchers to build cutting-edge technology. Some of the breakthroughs in these areas are highlighted in the following subsections.

Deep learning methods, particularly CNNs, RNNs, and LSTMs, are extremely capable of processing multimedia content for solving complex problems. CNNs are well suited for processing images and video frames for image classification, object detection and recognition, segmentation, enhancement, and restoration. On the other hand, RNNs and LSTMs are designed to process sequential data such as video event recognition, speech recognition, and image to text translation. Both these architectures can also be used in combination to solve sophisticated problems. For instance, in Ullah et al. (2018), CNNs are used as feature extractor and LSTMs are then employed to recognize patterns in video frames for event/action recognition. Both may be trained separately on the intended task and then used in combination during the inference phase. We have witnessed great progress in these methods in recent years, and we hope to see improvements in the coming years as well.

3.7.1 Computer Vision

Computer vision has been one of the first fields which saw breakthrough advancements due to deep learning. It has enabled computers to perform image recognition, object detection, and image segmentation at superhuman levels. The first notable achievement was seen in the year 1989 when a convolutional neural network was employed to read hand-written digits from snail mail (LeCun et al. 1989). The system was very robust and accurate; however, besides that, no notable achievement was noticed till 2012. AlexNet won the ImageNet competition beating the other teams by a huge margin. Since then, it grabbed the attention of researchers all over the world and rapid progress has been witnessed in the following years.

In a recent ImageNet competition, CNN with residual connections achieved human-level accuracy at image classification and even surpassed it when multiple networks were used in the ensemble (He et al. 2016). The tremendous improvement in computer vision has only been due to the efforts of researchers in the last 5–6 years which has made it possible to be used in critical application like self-driving cars and disease detection. Computer vision powered by deep learning has achieved expert-level accuracy in detection of skin cancer (Esteva et al. 2017), diagnosis in chest X-rays (Rajpurkar et al. 2017), and detection of diseases using various scans (Litjens et al. 2017). Besides these, it has also enabled computers to colorize black-and-white photos (Zhang et al. 2016), generate a high-resolution image from low-resolution input (Dong et al. 2016), pose estimation in real time (Jain et al. 2014), and generate captions for images and videos (Xu et al. 2015). All this progress is driven by CNNs which are very powerful architectures for computer vision.

3.7.2 Information Retrieval

The purpose of information retrieval (IR) is to allow people to access relevant information in the acceptable format in an efficient manner. The information may be in the form of images, audio, or text. In any case, machine learning and deep learning play a vital role in realizing the objective of IR. A typical framework for IR is shown in Fig. 3.5. A user issues a query in the form of text, image, or even audio clip. The IR system then performs a search in the database and retrieves the most relevant information in accordance with the query using machine learning to determine content relevant or similarity.

Text-based information retrieval systems are the simplest since the content matching process is very straightforward. However, IR systems really face challenge when dealing with highly complex data like images, videos, and audio. In these cases, it is essential to obtain an effective representation of the contents which will facilitate comparison in an efficient manner. For image retrieval, researchers previously relied on hand-engineered feature extraction algorithms like bag-of-words (BoW) (Yang et al. 2012), scale invariant features transform (SIFT) (Lowe 2004), and GIST

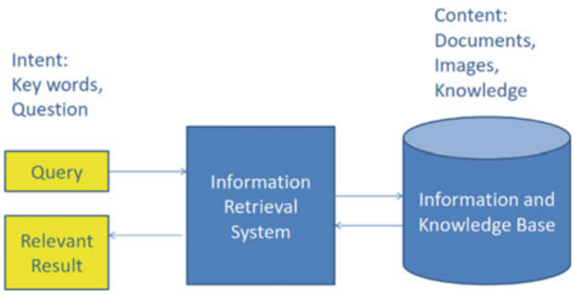

Fig. 3.5 Typical framework of an IR system

(Oliva and Torralba 2001) among many others. However, with the rise of deep learning methods, features automatically learned by a CNN have exhibited tremendous representation capability and have surpassed all handcrafted methods. Hence, convolutional features are becoming very popular with image retrieval experts. Neuronal activations from the deeper convolutional or fully connected layers are used as features to represent images in a content-based image retrieval (CBIR) system. The feature vectors of image pairs are then directly compared using basic Euclidean distance to determine their similarity.

Keeping in view the growing sizes of image and video datasets, efficient retrieval of relevant information has also become an important issue. For this purpose, hash-based representation schemes are developed, in which the image represented by a short hash code in Hamming space such that semantically relevant images lie close to one another (Ahmad et al. 2017a). In such a case, it becomes possible to retrieve relevant images without having to exhaustively search the entire database. This characteristic makes it very useful for dealing with very large datasets. Several supervised and unsupervised methods have been developed for generating hash codes for images. Recently, Ahmad et al. introduced an efficient method to transform high dimensional features to compact binary codes using Fourier decomposition which does not require any training (Ahmad et al. 2018b).

3.7.3 Natural Language Processing

Natural language processing (NLP) is considered as one of the most important technologies of information age. Understanding complex language utterances is a highly crucial component of artificial intelligence. There are widespread applications of NLP because people communicate almost everything in a language: Web search, advertisement, emails, customer service, language translation, radiology reports, etc. NLP applications are powered by machine learning. Recently, deep learning approaches have obtained state-of-the-art performance across many different tasks including NLP. These language models can often be trained with a single end-to-end model and do not require traditional, task-specific feature engineering. CNNs and RNNs have been employed to understand textual data in both short and long forms, ranging from sentiment analysis and opinion mining to text summarization and document recognition and language translation (Collobert and Weston 2008).

3.7.4 Multitask Learning

In machine learning (ML), we generally train a single model or a collection of models to perform our desired task. We can also try fine-tuning the model parameters and tweak these models until the desired performance is reached, or it no longer increases. Although acceptable performance can be achieved by being totally focused on a single task, sometimes we ignore information that might help us do even better on the metric we care about. Specifically, this kind of information comes from the training signals of other related tasks. By sharing representations among these related tasks, we can achieve better generalization for our model on our intended task. This approach is known as multitask learning (MTL).

MTL has been used successfully across several applications of machine learning, from NLP (Collobert and Weston 2008) and speech recognition (Deng et al. 2013) to computer vision (Girshick 2015) and drug discovery (Ramsundar et al. 2015). MTL is also referred to as joint learning, learning to learn, and learning with auxiliary tasks. Generally, when you are optimizing more than one loss function, you are effectively doing MTL. In such scenarios, it helps to think about what you are trying to do explicitly in terms of MTL and to draw insights from it.

Even if someone optimizing a loss, chances are there in an auxiliary task which will further assist to improve upon the main task. Rich Caruana (Caruana 1997) summarizes the goal of MTL succinctly as: "MTL improves generalization by leveraging the domain-specific information contained in the training signals of related tasks."

There are two methods for MTL, namely "hard parameter sharing" and "soft parameter sharing." Hard parameter sharing is generally applied by sharing the hidden layers between all tasks, while keeping several task-specific output layers as shown in Fig. 3.6. This approach greatly reduces the risk of overfitting.

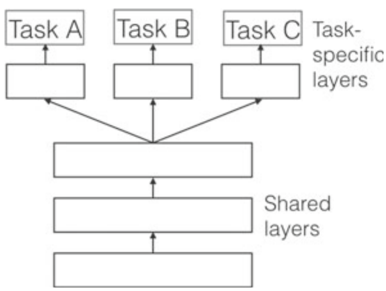

Fig. 3.6 MTL with hard parameter sharing

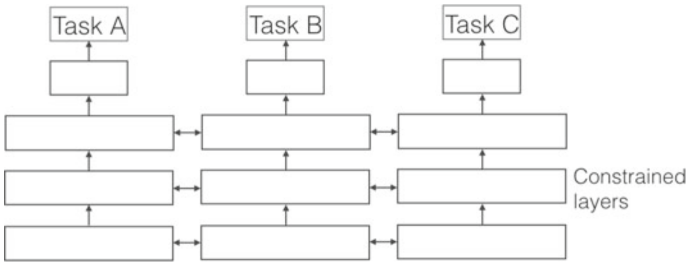

Fig. 3.7 MTL with soft parameter sharing

In soft parameter sharing on the other hand (shown in Fig. 3.6), each task has its own model with its own parameters. The distance between the parameters of the model is then regularized in order to encourage the parameters to be similar. For instance, (Duong et al. 2015) use the $\ell 2$ norm for regularization, while Yang and Hospedales (2016) use the trace norm. The constraints used for soft parameter sharing in DNNs have been greatly inspired by regularization techniques for MTL that have been developed for other models (Fig. 3.7).

References

Ahmad J, Muhammad K, Baik SW (2017a) Medical image retrieval with compact binary codes generated in frequency domain using highly reactive convolutional features. J Med Syst 42:24. https://doi.org/10.1007/s10916-017-0875-4

Ahmad J, Sajjad M, Mehmood I, Baik SW (2017b) SiNC: saliency-injected neural codes for representation and efficient retrieval of medical radiographs. PloS One 12:e0181707

Ahmad J, Muhammad K, Bakshi S, Baik SW (2018a) Object-oriented convolutional features for fine-grained image retrieval in large surveillance datasets. Future Gener Comput Syst 81:314–330. https://doi.org/10.1016/j.future.2017.11.002

Ahmad J, Muhammad K, Lloret J, Baik SW (2018b) Efficient conversion of deep features to compact binary codes using fourier decomposition for multimedia Big Data. IEEE Trans Ind Inf

Badshah AM, Ahmad J, Rahim N, Baik SW (2017) Speech emotion recognition from spectrograms with deep convolutional neural network. In: 2017 international conference on platform technology and service (PlatCon), 2017. IEEE, pp 1–5

Caruana R (1997) Multitask learning. Mach Learn 28:41–75

Collobert R, 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, 2008. ACM, pp 160–167

Deng L, Hinton G, Kingsbury B (2013) New types of deep neural network learning for speech recognition and related applications: an overview. In: 2013 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2013. IEEE, pp 8599–8603

Dong C, Loy CC, He K, Tang X (2016) Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Mach Intell 38:295–307

Duong L, Cohn T, Bird S, Cook P (2015) Low resource dependency parsing: cross-lingual parameter sharing in a neural network parser. In: Proceedings of the 53rd annual meeting of the associa-

tion for computational linguistics and the 7th international joint conference on natural language processing, vol 2 (Short Papers), 2015, pp 845–850
Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542:115
Girshick R (2015) Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision, 2015, pp 1440–1448
Glorot X, Bordes A, Bengio Y (2011) Domain adaptation for large-scale sentiment classification: a deep learning approach. In: Proceedings of the 28th international conference on machine learning (ICML-11), 2011, pp 513–520
He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp 770–778
Hinton G et al (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process Mag 29:82–97
Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313:504–507
Jain A, Tompson J, LeCun Y, Bregler C (2014) Modeep: a deep learning framework using motion features for human pose estimation. In: Asian conference on computer vision, 2014. Springer, pp 302–315
Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Paper presented at the proceedings of the 25th international conference on neural information processing systems, vol 1, Lake Tahoe, Nevada
LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1989) Backpropagation applied to handwritten zip code recognition. Neural Comput 1:541–551
LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444
Litjens G et al (2017) A survey on deep learning in medical image analysis. Med Image Anal 42:60–88
Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60:91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94
Lu N, Li T, Ren X, Miao H (2017) A deep learning scheme for motor imagery classification based on restricted boltzmann machines. IEEE Trans Neural Syst Rehabilitation Eng 25:566–576
Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comput Vision 42:145–175
Rajpurkar P et al (2017) Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:171105225
Ramsundar B, Kearnes S, Riley P, Webster D, Konerding D, Pande V (2015) Massively multitask networks for drug discovery. arXiv preprint arXiv:150202072
Ullah A, Ahmad J, Muhammad K, Sajjad M, Baik SW (2018) Action recognition in video sequences using deep Bi-directional LSTM with CNN features IEEE. Access 6:1155–1166
Xu K et al (2015) Show, attend and tell: neural image caption generation with visual attention. In: International conference on machine learning, 2015, pp 2048–2057
Yang W, Lu Z, Yu M, Huang M, Feng Q, Chen W (2012) Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single-and multiphase contrast-enhanced CT images. J Digital Imaging 25:708–719. https://doi.org/10.1007/s10278-012-9495-1
Yang Y, Hospedales TM (2016) Trace norm regularised deep multi-task learning. arXiv preprint arXiv:160604038
Zaremba W, Sutskever I, Vinyals O (2014) Recurrent neural network regularization. arXiv preprint arXiv:14092329
Zhang R, Isola P, Efros AA (2016) Colorful image colorization. In: European conference on computer vision, 2016. Springer, pp 649–666

Chapter 4
Integration of Big Data and Deep Learning

Muhammad Talha, Shaukat Ali, Sajid Shah, Fiaz Gul Khan and Javed Iqbal

Abstract The traditional algorithms of artificial intelligence and neural networks have many limitations to process big data in real time. Therefore, the researchers introduce the concept of deep learning to address the aforementioned challenge. However, big data analytics required a process consists of various steps where in each step an algorithm or a bunch of algorithm can be used. This chapter explains the role of machine learning in processing big data to meet various applications and users' demands in real time. Similarly, various techniques of deep learning are studied to show how they can be used to address various challenges and issues of big data. Similarly, other similar techniques such as transfer learning are also discussed to support the study of deep learning.

List of Acronyms

CNN	Convolutional neural network
DBN	Deep belief network
GPU	Graphical processing unit
RBM	Restricted Boltzmann machine
DSN	Deep stacking network
RFID	Radio frequency identification

4.1 Machine Learning in Big Data Analytics

Companies are collecting a huge amount of data at different pace, volume, and format. The volume of data is not that much important, but what really matters is the use of this data. In today's digital world, big data has a very significant role when it comes to customer data. But this data can be of no use if it is not paired with machine learning systems having high computational capability. Machine learning systems can give a real insight into the customer data and can be used in more efficient way according

M. Khan et al., *Deep Learning: Convergence to Big Data Analytics*, SpringerBriefs in Computer Science, https://doi.org/10.1007/978-981-13-3459-7_4

to the needs of the organization. Big data along with machine learning can really boost the organization business and can gear up long-time business importance.

The distinctive features of big data such as huge volume, different varieties, high speed, complex, unstructured, and inaccurate have really challenged the traditional data mining and statistical techniques, which are primarily developed for small datasets. Machine learning techniques are favored by data scientists to explore the patterns and structures hidden in big data and to extract more useful information out of it. Machine learning is categorized into supervised and unsupervised learning, already discussed in Chap. 1. Popular machine learning approaches are support vector machines, classification trees, Naïve Bayesian, and K-nearest neighbor.

4.1.1 Machine Learning and Big Data Applications

Machine learning is widely used in different real-world applications in different disciplines. The use of machine learning is there where we have a high volume of structured and unstructured data. Recently it has attracted scientists from different domains such as machine learning which is used in plant sciences to analyze large datasets (Ma et al. 2014). Data in biomedical and healthcare is generated on daily basis, and it is very important to analyze with maximum accuracy and minimum response time (Chen et al. 2017). Machine learning along with big data is used in healthcare for improving patient care and monitoring, early disease detection, and assisting doctor to forecast patient health issues. Automotive, financial services, intelligent transportation systems, national security, computer vision, and many others are using machine learning to extract data of their requirement (Data Science 2018).

In this digital world where data is generated in such a huge volume with different varieties and veracity, it becomes a top priority for organizations to handle it efficiently. The impact is that companies are restructuring infrastructures and shifting toward big data to increase automation and the use of smart devices to enhance the productivity and to deliver services to the customer in the best possible way. Machine learning systems with high computational and storage capacity along with intelligence can offer such services, and big companies have already integrated machine learning and big data.

4.2 Efficient Deep Learning Algorithms in Big Data Analytics

As earlier discussed, deep learning is the subfield of machine learning, and it is one of the hottest topics that is adopted in almost every field where ever big data is involved. It is expected that by 2020, the amount of data generated over the Internet will cross 35 trillion GBs. Now, one can imagine that what kind of challenges it will bring when it comes to extract useful information according to the need of organizations. Deep learning is the answer to many of the questions that are associated with this bulk of data. Deep learning automatically learns patterns and structures hidden in the raw data using machine learning techniques. Due to its features, DL has not only attracted researchers from different domains but also got popularity in the industry, and companies such as Facebook, Apple, Google, and YouTube are pushing deep learning toward their services and products (Chen and Lin 2014). For instance, Apple's Siri (Efrati 2017) is the virtual personal assistant, Google using it in Google translate, Google street view, image search engine, and voice recognition (Jones 2014).

Some of the deep learning models and algorithms are deep belief networks, recursive neural networks, convolutional neural networks, convolutional deep belief networks, and deep Boltzmann machines. However, the two commonly used architectures in deep learning are deep belief network (DBN) and convolutional neural network (CNN) (Chen and Lin 2014).

Deep belief networks have the potential to learn the representation of features using structured and unstructured data with the help of supervised and unsupervised techniques. It consists of input, hidden, and output layer. The restricted Boltzmann machine (RBM) uses DBN to construct a model that consists of two layers that are fully connected to each other (Raina et al. 2009). In the literature, DBN model is used by many researchers to efficiently and accurately process big data. M. A. Raina proposed a graphical processing unit (GPU)-based model using stacked RBM in parallel to handle large volume of data with minimized process time (Raina et al. 2009). The power of deep learning is that it can train and handle millions of parameters at a time. Deep stacking network (DSN) is proposed by D. Y. L. Deng, which consists of a single hidden layer having many specialized neural networks (modules). These modules are parallelized with inputs, and the output of each module constructs DSN (Deng et al. 2012). Tensor DSN is the modified deep architecture proposed for parallel computing consists of CPU clusters (Hutchinson and Yu 2013).

The training process can be speeded up by maximizing computing power, which has great importance in big data using deep learning. Multiple processing cores are used, where each core handles a subset of data. Another approach splits the hidden and visible units in n machines to speed up the process. FPGA-based implementation is also used for large-scale deep learning. Convolutional neural network is another most commonly used architecture, in which deep learning methods are locally connected. It is composed of feature map and classification layers in a hierarchical manner. The layer that takes data from input layer is known as convolutional layer, responsible for convolutional operations (Deng and Yu 2014). The output of this layer is forwarded to sampling layer that reduces size of upcoming layers. CNN is implemented on number of cores having GPU implementation to support the large number of layers. The CNN architecture has been widely used in different domains such as computer vision and image and speech recognition. Two GPUs are proposed by A. Krizhevsky with five convolution and three classification layers to achieve high-speed processing (Krizhevsky 2012). Authors proposed that half of the layers will be processed by one GPU and half by another to distribute the load of processing without affecting host memory. CNN has been used in different domains such as fire detection, video surveillance, disaster management, computer vision.

4.3 From Machine to Deep Learning: A Comparative Approach

Machine intelligence is a field of artificial intelligence and refers to the intelligence exhibited by computers through programming. However, when we are moving toward big data, traditional machine learning techniques are turned to deep learning techniques, which are more powerful, easy to implement, and efficient to deal with huge amounts of data originated from recent data-driven businesses. Deep learning, which is a subset of the machine learning approach, is a new and more complicated way of analyzing big data. Deep learning allows us to solve those problems which were impossible to address before the introduction of deep learning. So, deep learning is an approach for implementing machine learning techniques. There are some tasks that machines can perform it in a better way than human beings; for example, in image classification, the computer shows better results than human. In 2015, the ImageNet winning entry, the ResNet, has given better accuracy than human level in with less error rate compare to traditional methods.

I would like to explain the comparison between traditional machine learning and deep learning through an example in which the system recognizes an animal; the system has to identify the given image of an animal. To solve this problem through machine learning approach, all features like if the animal has whiskers or not, similarly, the domain expert will define all the important features. In contrast to machine learning, the deep learning repeatedly selects that features which are required for categorization of the animals. Following are the important factors that could be used to compare the traditional machine learning and deep learning:

4.3.1 Performance on Data Size

The performance of deep learning and machine learning is dependent on the size of dataset. When the volume of data increases, the deep learning algorithms execute quickly than traditional machine learning algorithms. However, in the case of a small dataset, the conventional machine learning approach shows better results than deep learning algorithms.

4.3.2 Hardware Requirements

Deep learning algorithm greatly depends on powerful machines; on the other hand, traditional machine learning algorithms can also work on low powered devices. It is because the deep learning algorithms essentially perform a large size matrix multiplication.

4.3.3 Feature Selection

In machine learning, almost all the features are identified by a domain expert and then hard coded as per the environment and data type. While in deep learning, algorithms perform the learning process of high-level features from source data. Therefore, deep learning decreases the task of programmer and new features are extracted for such problems. For example, CNN tries to learn low-level features like edges and lines in early stages then parts of faces of people and then high-level representation of a face.

4.3.4 Problem-Solving Approach

When we are solving some problem using machine learning algorithm, it normally divides the problem into multiple parts. It addresses these parts individually and combines them when solving the problem and gets the result. Deep learning, in contrast, solves the problem from end to end.

4.3.5 Execution Time

Normally, a deep learning algorithm takes more time to train it in contrast traditional machine learning which takes lesser time to train. The machine learning takes training time ranging from some seconds to a few hours. However, in the testing stage, the deep learning approach takes much lesser time than machine learning.

4.4 Applications of Deep and Transfer Learning in Big Data

Transfer learning is comparatively a new approach for the improvement of data learning by using the acquired knowledge from different known class-labeled data tasks and domains (Yang 2008). The transfer learning is to use the existing knowledge learned from a task where a lot of known class-labeled data is available to those data where a little known class-labeled data is available. The traditional machine learning model generalizes to unknown data based on learned patterns from the training set. While in transfer learning, the generalization process is started from those patterns that have been learned from other different tasks. The machine learning approaches require the training dataset, and test dataset is supposed to have the same feature space and the equal distribution. In contrast, the transfer learning approach allows the tasks and distribution of the dataset used for training and testing purposes may be different. It is generally used in those situations when we have the training data that are insufficient for an appropriate data modeling. The knowledge is transferred from the related training data from some other tasks to enrich the data features of the target task. In transfer learning, more data characteristics are integrated, more improved will be the learning results (Yang and Chu 2015).

Deep learning has its great success and achievement in many areas, for example, image recognition and analysis, speech recognition, and text mining. It can be used both as a supervised and unsupervised learning strategy to learn the multilevel features and representations for the tasks of classification and pattern recognition. In the recent era, the sensor networks and the advancement of communication technologies have generated a huge amount of data. However, this big data offers great prospects for different areas like industrial control, e-commerce, and smart medical. Along with these opportunities, it has many issues on data mining and information processing due to its characteristics of large volume, variety, velocity, and veracity. The deep learning plays a significant role in big data analytics (Zhang et al. 2018). The deep learning techniques can be used to parallelize different data-driven applications to achieve optimum performance. For example, Hernández et al. (2017) used a benchmark of 15 Spark applications on the Grid5000 test bed and was claimed a 51% performance improvement while using the recommended parallelism settings.

4.4.1 Healthcare

Deep learning can work as an assistant to medical practitioners and researchers to extract the hidden relationships in the data related to healthcare and to serve it in a better way. Deep learning can be used by doctors to analyze many of the diseases accurately and can take help from deep learning to treat them in a better way.

In drug discovery, the deep learning can help to discover a new or develop the existing medicines. The deep learning techniques can analyze the patient's medical history and suggests the best treatment for them. Moreover, the patient data can be used for gaining insights from patient symptoms and tests to predict the future patterns.

The deep learning techniques can be applied to medical imaging like MRI scans, CT scans, and ECG to diagnose alarming diseases, for example, heart disease, cancer, and brain tumor. It helps the doctors to analyze the dangerous diseases in a better way to provide the patients with the best treatment. Deep learning can be used to detect the cancer-type diseases at an early stage.

In the patient insurance schemes, frauds can be detected using deep learning approaches by analyzing the medical insurance and fraud claims data. It can predict such fraud claims that may occur in the future. The deep learning can also be used for the promotion of the insurance industry in the medical field by analyzing the patients' data.

4.4.2 Finance

Finance is one of the computationally intensive fields which requires a lot of computation when we are dealing with a financial model. It is because the economic domain is complex and nonlinear with a large number of factors that influences other factors. Deep learning can be used in such situations. These techniques can be used in risk management, pricing, and even for the prediction of the future trades.

Deep learning approach can be used for better customer services. It can be used for the categorization of customers based on the loyalty to give them better services and facilities. In financial matters, fraud detection is one of the most challenging and vital tasks. Big data can be used to analyze it using machine learning techniques and detect the frauds.

4.5 Deep Learning Challenges in Big Data

Processing big data is a challenging job, and many researchers and data scientist are still working in the same field. The data around us is generated with high speed, and its volume is getting increasing every day. Therefore, we need sophisticated algorithms to target the big data before it goes out of the reach of the existing algorithms. Before going to discuss the big data challenges in detail, we will present a comprehensive analysis of how data is generated from various sources. These sources include but not limited to exisiting datasets and databases containing structured and unstructured data.

4.5.1 Internet of Things (IoT) Data

As we know in the world economic forum 2011, the data is titled as the new oil of the world. Therefore, we need proper and efficient data analytics to care about various dimensions of big data. However, with the introduction of IoT concepts, the nature of data is entirely changed in the size, shape and requires statistical methods to address the analytics of data. Every day, millions of GB data is generated due to the millions of connected devices over the Internet. These devices are connected through various technologies such as sensor network, radio frequency identification (RFID), Bluetooth, and Wi-Fi. One of the major problems is to use sophisticated techniques to handle the data during the communication over these technologies. Similarly, the IoT is still in its initial phases and we need standards and laws to regulate the use of IoT.

4.5.2 Enterprise Data

A recent study has suggested that the enterprise data is the world's main source of generating big data. The most important sources include the Web sites and social networking Web sites operating over the Internet such as Facebook and Google. However, to deal with such huge amount of data generated by these devices everyday needs proper energy management. The energy management can be carried out using the green technologies. However, still we need further assessments and methods to come up with the proper methods and techniques for green computing.

4.5.3 Medical and Biomedical Data

The biomedical data is generating with several properties such as heterogeneous structures, large amount, and biological concepts. Dealing with such data can produce beneficial results both for medical research and dealing with new diseases. However, it is still a challenging area and the researchers and data scientists are working to use proper tools and techniques to come up with better results and outcomes.

Apart from these generating sources and its challenges, the four Vs structure of the big data is widely studied and addressed at various forums. Figure 4.1 shows the challenges present in the form of four Vs.

Similarly, traditional learning algorithms are not designed to handle continuous stream of data. Therefore, such properties of big data lead to another dimension of big data called real-time processing. Processing off-line data with Hadoop alongside MapReduce structure is somehow possible. However, online data or the data generated cannot be processed in real time using the Hadoop ecosystem.

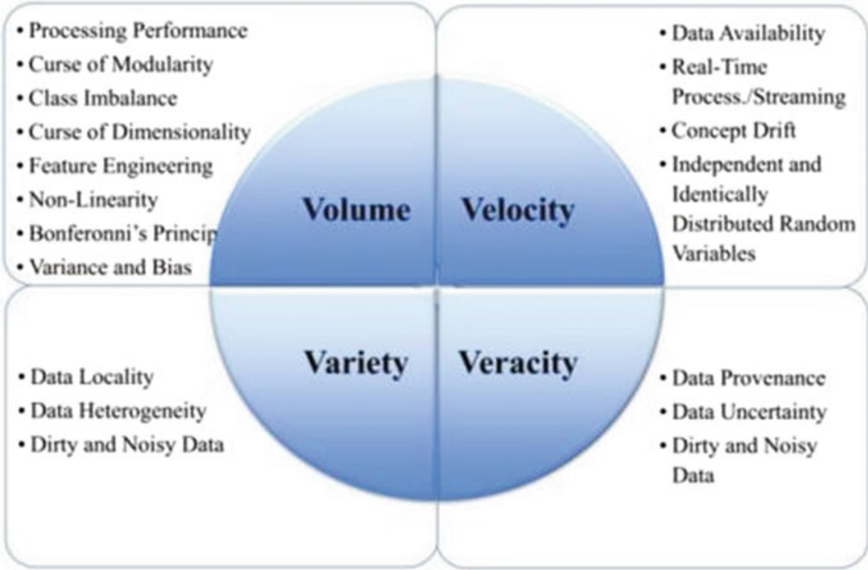

Fig. 4.1 Challenges of big data present in the form of four Vs

References

Chen X-W, Lin X (2014) Big data deep learning: challenges and perspectives. IEEE Access 2:514–525

Chen M, Hao Y et al (2017) Disease pre-diction by machine learning over big data from healthcare com-munities. IEEE Access 5:8869–8879

Deng L, Yu D (2014) Deep learning: methods and applications. Found Trends Sig Process 7:197–387

Deng L, Yu D, Platt J (2012) Scalable stacking and learning for building deep architectures. In: Paper presented in IEEE international conference on acoustics, speech and signal processing, Kyoto Japan, 25–30 Mar 2012

Efrati A (2017) How deep learning works at apple, beyond. https://www.theinformation.com/How-Deep-Learning-Works-at-Apple-Beyond

Hernández AB, Perez MS et al (2017) Using machine learning to optimize parallelism in big data applications. Future Gener Comput Syst 86:1076–1092

Hutchinson LDB, Yu D (2013) Tensor deep stacking networks. IEEE Trans Pattern Anal Mach Intell 35:1944–1957

Jones N (2014) Computer science: the learning machines. Nature 505:146–148

Krizhevsky A, Sutskever I et al (2012) Imagenet classification with deep convolutional neural networks. In: Paper presented in advances in neural information processing systems, Lake Tahoe, Nevada, USA 3–8 Dec 2012

Ma C, Zhang HH et al (2014) Machine learning for big data analytics in plants. Trends Plant Sci 19:798–808

Raina R, Madhavan A, Ng AY (2009) Large-scale deep unsupervised learning using graphics processors. In: Paper presented in proceedings of the 26th annual international conference on machine learning, Montreal, Quebec, Canada, 14–18 June 2009

Towards Data Science (2018) https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d. Accessed 14 Sept 2018

Yang Q (2008) An introduction to transfer learning. ADMA

Yang L, Chu Y et al (2015) Transfer learning over big data. In: 10th international conference on digital information management (ICDIM). IEEE

Zhang Q, Yang LT et al (2018) A survey on deep learning for big data. Inf Fusion 42:146–157

Chapter 5
Future of Big Data and Deep Learning for Wireless Body Area Networks

Fasee Ullah, Ihtesham Ul Islam, Abdul Hanan Abdullah and Atif Khan

Abstract Deep learning is an innovative set of algorithms in machine learning and requires minimum efforts of human engineering in extraction of features from data. It has the ability to find the optimum set of parameters for the network layers using a back-propagation algorithm, thereby modeling intricate structures in the data distribution. Further, deep learning architectures have resulted in tremendous performance on most recent machine learning challenges included working with sequential data such as text and time series data. In this connection, big data technology is an asset for modern businesses and is useful if powered by intelligent automation. Big data involves huge datasets that can be analyzed by machine learning such as deep learning algorithms to find insightful patterns and trends. With modern-day machine learning and big data technology, organizations can drive its long-term business value far more successful than ever before. Potential real-world applications of big data are not limited to healthcare, retail, financial services, and the automotive industry. In this way, the deep learning can have a great impact on analyzing the patient's data generated from wireless body area networks (WBANs). WBAN is the emerging technology in healthcare to assist in monitoring of vital signs of patients using biomedical sensors. The monitored data is transmitted to the medical doctor for an optimal treatment in a life-threatening situation. At the end of this book, open research issues in WBAN and big data have discussed.

List of Acronyms

BMS	Biomedical sensor
CAP	Contention-access period
CNN	Convolutional neural networks
CEP	Complex event processing
CGOC	Compliance, Governance and Oversight Council
CFP	Contention-free period
CS	Conventional server
CSMA/CA	Carrier-sense multiple access with collision avoidance

M. Khan et al., *Deep Learning: Convergence to Big Data Analytics*, SpringerBriefs in Computer Science, https://doi.org/10.1007/978-981-13-3459-7_5

DNN	Deep neural network
EAP	Exclusive access phase
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyography
IEEE	Institute of Electrical and Electronics Engineers
IP	Inactive period
GPU	Graphics processing unit
GSM	Global system for mobile
GST	Guaranteed time slot
HDFS	Hadoop Distributed File Systems
LOS	Line-of-sight
LSTM	Long short-term memory
MLP	Multilayer perceptron
MAC	Medium access control
NLOS	Non-line-of-sight
PHY	Physical layer
QoS	Quality of service
RAP	Random-access phase
RNN	Recurrent neural network
SPO2	Peripheral capillary oxygen saturation
TDMA	Time-division medium access
VC	Virtualized cloudlet
WBAN	Wireless body area networks
WHO	World Health Organization
WSN	Wireless sensor network
TG6	Task Group 6

5.1 Introduction

Machine learning is a modern-day technology that enables systems to learn from experience using statistical techniques, where it is difficult to explicitly program the computing tasks involved. It leverages data to create intelligent programs and has tremendous applications in almost any industry not limited to healthcare, banking, finance, agriculture, manufacturing, and automation. From content filtering on social networks and self-driving cars, it is gradually employed to handheld devices and consumer products. Specific goal of a machine learning system can be identifying biological cells in a microscopic image, converting speech to commands, translating text to a different language, or recommending the next movie to watch.

Recently, these applications are revolutionized by a class of algorithms based on neural networks, called deep learning. Advanced tools and techniques have dramatically transformed the conventional neural network algorithms to the point where they can outperform humans. Simple neural network design allowed 2–3 layers,

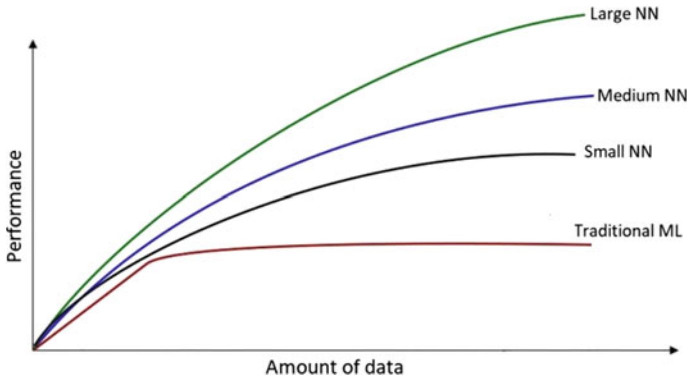

Fig. 5.1 Performance relationship of deep networks with the increase in amount of data original slide by Professor Andrew Ng

while deep networks may contain hundreds of layers. The successful design and applicability of a deep network is possible due to three main technologies. First, the availability of huge label datasets allows it to capture the actual distribution of the patterns. Second, together with high-performance GPUs and increased memory, it has made it possible to reduce training time of deep networks with huge data from months to hours. Last but not least, pretrained models can be retrained with smaller data sizes for new tasks, a technique known as transfer learning, which results in saving training time and effort yet not compromising performance. Figure 5.1 shows the performance relationship of deep networks with the increasing amount of data. It is clear that traditional machine learning can hardly benefit from the increase in data size.

Deep learning learns to extract features directly from images, text, and sound, at multiple layers of the networks, where complex features are defined across the layered hierarchy in terms of simple low-level features. It then also learns a function to map the features into a desired output and achieves higher accuracy with more data, entitling deep learning as an end-to-end learning algorithm. This is contrary to other machine learning algorithms that learn only a function that maps input features to a desired output. For the reason, machine learning workflow needs an explicit feature engineering or feature extraction step done by human engineers, resulting in features termed as handcrafted features. See Fig. 5.2.

The rest of the chapter is constructed as follows: Sect. 5.2 presents deep learning frameworks based on the feed-forward network model. Future of deep learning in big data is discussed in Sect. 5.3. Further, introduction of wireless body area network is presented in Sect. 5.4. Section 5.5 presents the existing applications and future applications of wireless body area network. Section 5.6 presents the existing challenges in routing protocols of wireless body area network. Working of superframe structures of IEEE 802.15.4 MAC and IEEE 802.15.6 MAC along with research challenges has been presented in Sect. 5.7. Further, introduction to big data is described in Sect. 5.8, and the applications of big data are presented in Sect. 5.9. Open research issues of WBAN and big data are presented in Sects. 5.10 and 5.11, respectively.

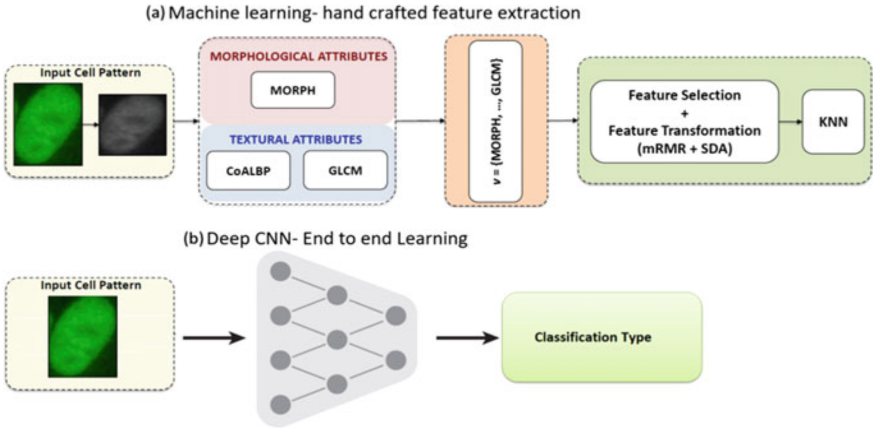

Fig. 5.2 **a** (Di Cataldo et al. 2014; Ul-Islam 2014) Traditional machine learning with feature engineering by a domain expert versus end-to-end deep convolution learning **b** Deep CNN-end to end learning

5.2 Feed-Forward Network Model

A deep neural network loosely mimics the human mind. It has layers with nodes or neurons which are interconnected from the previous layers that act as input activations. The layered construction utilizes the unknown data distribution and models to capture nonlinearity, resulting in better representation. It consists in an input layer, hidden layer(s), and an output layer(s). Modern-day neural nets can be classified into a number of types based on their design and application. Among are multilayer perceptions (MLPs), deep convolution neural nets (Lecun and Bottou, n.d.), recurrent neural nets (Schmidhuber 1997; El Hihi and Bengio 1996), and many others like autoencoders (El Hihi and Bengio 1996), generative adversarial networks (El Hihi and Bengio 1996), and residual networks (He et al. 2016). Figure 5.3 shows MLP with one input and output layers and with two hidden layers.

In a forward pass, the input layer accepts inputs presented to the model; subsequently, each layer linearly combines the weighted output from the previous layer, and at each node, an activation function is applied. Input x_i is the applied at the input layer, where at each node in the hidden layer the activation a is calculated as the weighted combination of the input variables, $a = \sigma(z)$ $and\ where\ z = w^T x + b$, from the previous layer in the network.

Activation function characterizes the output of a neuron and introduces nonlinear transformation making the network capable of learning complex tasks. On contrary, its absence would limit the network nodes with simple linear transformations and never be able to model complex structures. Many kind of activation functions are commonly used in practice like the hyperbolic tangent function $tanh$, sigmoid function, ReLU (Taigman et al. 2014), and leaky ReLU functions, Fig. 5.4. Intuitively,

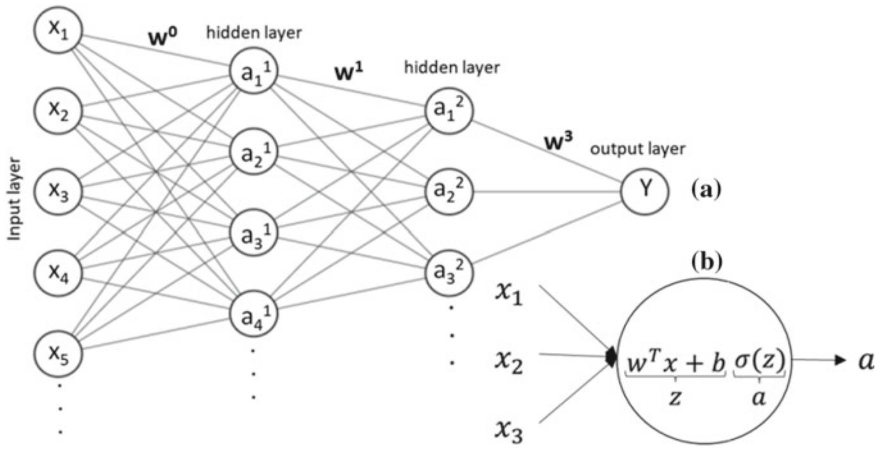

Fig. 5.3 **a** A feed-forward network (or MLP) with input layer, hidden layers, and output layer and **b** the computation at a single neuron with input and output activation

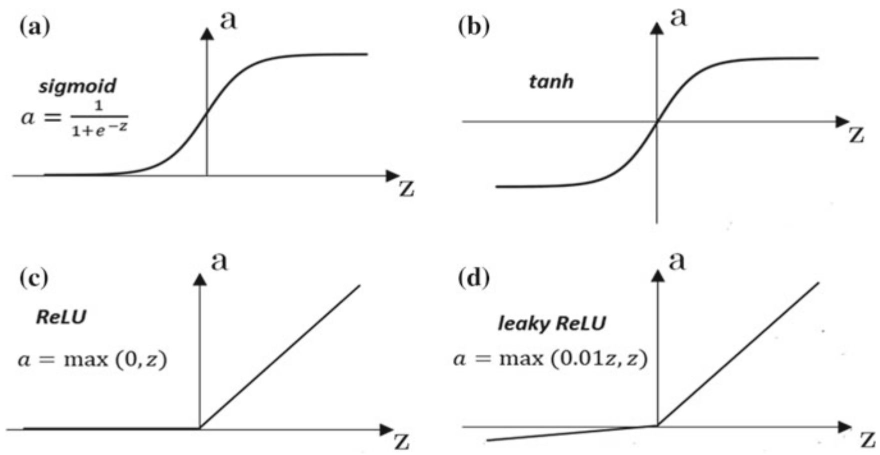

Fig. 5.4 Different activation functions are shown in **a** sigmoid, **b** tanh, **c** ReLU, and **d** leaky ReLU

each node in the network is responsible for detecting a particular feature. In case of $tanh$ function, a smaller value of z would give a higher gradient, suggesting further training of the node, and a larger z would seize such training. Calculation of these gradients and retraining of weights are part of a back-propagation step of the deep learning algorithm. A heuristic activation function ReLU has become quite popular, due to its fast computation; however, its updating of weights sometimes leaves a node as inactive. This problem is being resolved by utilizing, for example, a leaky ReLU (Maas and Hannun 2013) at the expense of additional computations. Figure 5.4 shows the different activation functions.

To make this all happen, a cost is always computed at the output layer, which is a measure of the difference between output value and actual ground truth value. A cost function is represented as $J(w, b) = \frac{1}{m}\sum_{i=n}^{m} L(y', y)$, where w are the weights, b is the bias, m is the number of training samples, y as actual and y' as the estimated output, and L is the loss function. The popular loss functions are cross-entropy loss, misclassification rate, or $L2$ loss also known as the mean squared error. The objective of the training process is to minimize this cost, using an optimization algorithm such as gradient descent algorithm which involves taking gradients of the activations and updating weights at network layers. This process is called back propagation. The final weights that minimize the loss function are considered to be the solution of the DNN model.

5.2.1 Deep Learning Frameworks

Among the types of deep neural nets is convolution neural network (CNN), which has attracted many researchers in the computer vision community (Multi-column deep neural network for traffic sign classification 2012; Taigman et al. 2014; Hadsell et al. 2014; Sainath et al. 2013). The algorithm works well where the inputs are images or if the data modalities are in the form of multiple arrays. Like MLP, CNN also contains an input layer, several hidden layers, and an output layer. The layers involve some operations on data termed as convolution, pooling, and the ReLU function. Convolution uses filters on image or multiarray to highlight certain features. Pooling involves nonlinear downsampling or reducing the parameters of the network. ReLU activation is applied for the purpose of faster training. Some top CNN architectures are AlexNet (Krizhevsky et al. 2012), VGG (Karen Simonyan 2014), Inception V3 (Szegedy et al. 2016), and ResNet (He et al. 2016), resulting in significant gains on the popular ImageNet dataset (Sutskever et al. 2011), with 1000 classes and 1.2 million labeled images. The trained models are useful to study, as they can be utilized to perform transfer learning, a technique that helps to reduce the training time on new tasks.

Whereas CNNs are winner algorithms for processing images and multiarray modalities like video, speech, and audio, recurrent neural net (RNN) has excelled on tasks that involve sequential data such as language modeling and translation, speech recognition, handwriting recognition, and other sequence problems. They can be used to identify the next character in a word (Sutskever et al. 2011) or the next word in a sentence or can be used for more complex task such as outputting the sentiment expressed in a paragraph. A sequence is simply a stream of data items, where the individual items are interdependent. For instance, the meaning of a sentence can be correctly understood once we put the entire workflow of conversation into context. Similarly, in stock market, a single tick will only tell the current price, but to model the movement and enable a buy–sell decision, more data readings are needed in the sequence. The construction of RNN is such that the hidden state of the previous time step and the input of the current time step are used to calculate the current output. Unlike the basic NN, RNN uses its internal memory to process somewhat arbitrary sequences of inputs. One of the famous algorithms is long short-term memory (LSTM) (Schmidhuber 1997) which is a widely class of RNNs.

5.3 Future of Deep Learning

The progress in deep learning and AI-based solutions will continue and is expected to accelerate in the coming future. Although this success is only seen in supervised learning-based solutions, the future holds reviving interests for the research community in unsupervised learning. This is plausible, as the human learning process is largely unsupervised, we model structures in the world by observation not by being taught by its names. To this end, potential future work in deep learning for computer vision can involve in combining different representations like CNN, RNN, and deep reinforcement learning. Initial advances in this regard have resulted in interesting applications like a computer learn to play video games. With increasingly massive data sizes and computational facilities, end-to-end deep learning-based algorithms and architectures are expected to see more successes in the near future.

5.4 Introduction to Wireless Body Area Networks

Wireless body area networks (WBANs) are the emerging technology and the most attractive field for the research community (Cavallari et al. 2014), academia (Quwaider and Jararweh 2015), and industry (Shu et al. 2015) to solve the health monitoring issues for patients in urban and rural areas. Further, the World Health Organization (WHO) (Latré et al. 2010a; Acampora et al. 2013; Murray et al. 2012) has issued various reports on the death rates that have been increased in millions annually due to various diseases. These diseases are cancer, heart attack, diabetic issues, stroke, respiratory, abnormality of blood sugar (Acampora et al. 2013). Further, WHO has categorized these death rates based on the income of countries which are low income, middle income, high income worldwide as shown in Fig. 5.5. Figure 5.5 depicts the increased number of death rates due to insufficient health resources to provide to patients on time (Latré et al. 2010a, b). The existing studies show that most of the deaths are home-based old aged people and also because of remotely monitoring of patients from remote areas which are away from routine health checkup. In addition, the existing hospitals cannot provide health services on time due to lack of manpower and resource-constraint environment of hospitals which are the costly practices. Thus, for long-term monitoring of health problems of patients and home-based old aged people, WBAN is the innovative technological and cheap solutions to

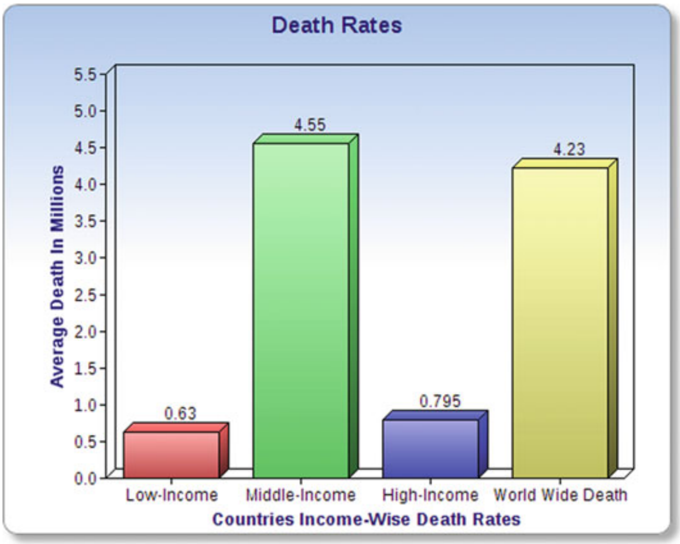

Fig. 5.5 WHO reports on average death rates in million

monitor vital signs of a patient. The monitored data is then forwarded toward medical doctors for an optimal treatment via GSM technology.

WBAN employs various biomedical sensors (BMSs) to monitor various vital signs of a patient. These BMSs are included to monitor respiratory rate, heartbeat rate, blood pressure, glucose level, temperature, ECG, EEG, EMG, and SPO2 (He et al. 2011) of patient. The monitored (sensory) data of a patient is transmitted to coordinator where all types of sensors are connected with coordinator as shown in Fig. 5.6. The coordinator is responsible for transmission of the sensory data to medical doctors. BMSs are connected with coordinator in the star or mesh topology based on the need of a patient. Figure 5.6 shows three methods of deployment of BMSs with a coordinator in the star topology. The first method is known as implantable, whereas different BMSs are inserted inside the patient's body to monitor various organs like lungs, kidney, and heart. The suitable example of the implantable sensors is capsule endoscope, as shown in Fig. 5.7a. Using wearable sensors is the second method to monitor vital signs like using of ECG, EEG, and EMG sensors which are sewed in the shirt of a patient or directly placed on the skin of a patient, as shown in Fig. 5.7b. The third method of deployment of sensors is different where different behavioral monitoring sensors are placed around the patient to monitor different physical activities. These physical activities are included monitoring of sleeping position and duration, postural movements of a patient like running, walking, dancing, and defective setting on sofa.

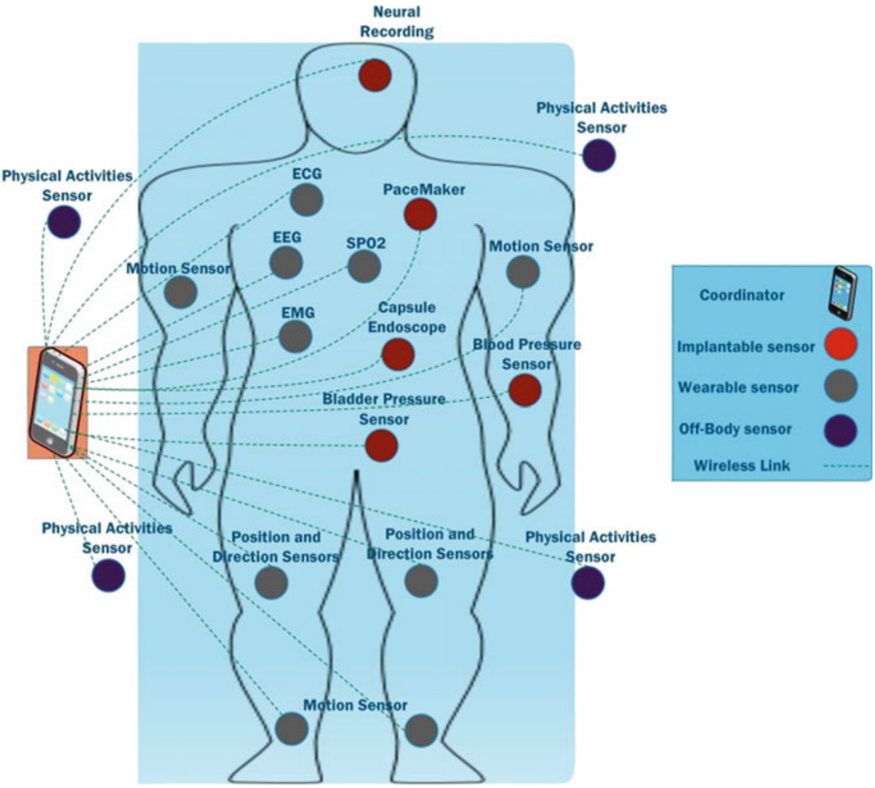

Fig. 5.6 Typical deployment of BMSs to monitor vital signs of patient

(a) Implantable BMS for Heart Monitoring in real time

(b) Wearable of BMSs

Fig. 5.7 Typical deployment of BMSs

5.5 Applications of Wireless Body Area Networks

Wireless body area network has wide applications in many fields including emergency services, consumer electronics, sports and fitness, lifestyle, defense, entertainment and gaming, personal healthcare, and medical (Applications of WBAN 2018), as shown in Fig. 5.8. In emergency services and defense, the deployed BMSs monitor vital signs of firefighters and soldiers, respectively, in their working environment. The aim of this monitoring is to inform the medical doctors if the person wounds or any emergency situation occurs. The sports and fitness, personal healthcare, and medical are together to monitor and maintain his/her health during sports activities, monitor home-based aged people health, and monitor different vital signs of serious patients in intensive care units (ICUs) and wards, respectively. The entertainment and gaming and consumer electronics are come in categories of enjoyable applications, whereas a person can open, forward, install, and delete game or songs, and similarly, a person can download, install, and remove an app, based on the mood, respectively. With these applications serving for humanity, it may be used widely.

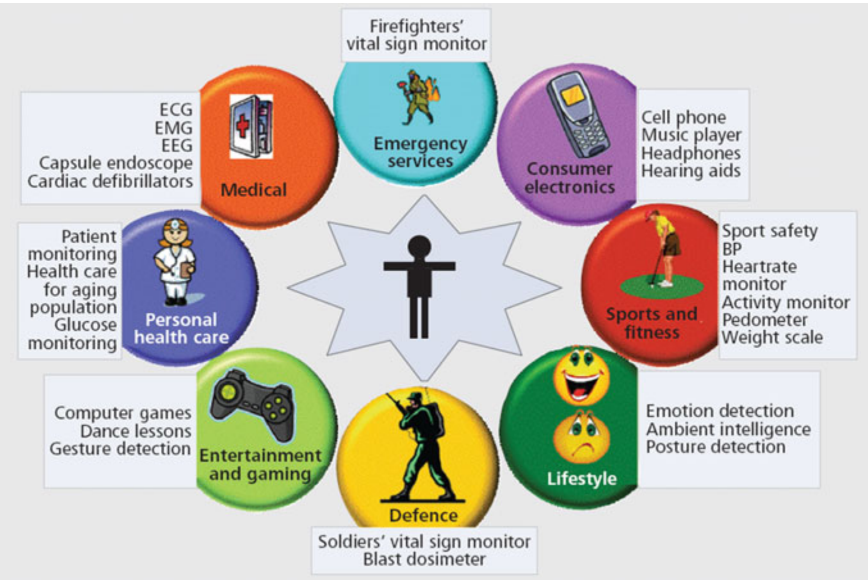

Fig. 5.8 Applications of wireless body area network (Applications of WBAN 2018)

5.5.1 Future Applications of Wireless Body Area Networks

The future of technological applications of WBAN is emerging toward driving assistance, electronic bill payment, security-based entrance to office, automatic operation in printing, and monitoring of health condition of patient (Future Applications of WBAN 2018), as shown in Fig. 5.9. In future, the driving person can adjust a seat of car if she feels uncomfortable with her driving seat. In this situation, she will apply force toward backside of seat for adjustment. In addition, the deployed BMSs will monitor different vital signs and will inform medical doctors if a person feels uncomfortable. In similar way, the person will not keep all transaction cards in his pocket but the embedded sensor in body will authenticate user and perform transactions. Furthermore, the outnumber of security guards will be reduced by company and the implanted sensors will authenticate users either he/she can enter to office or not. Printing of documents will be easy because bar code on the paper will recognize how much documents need to be printed. In this situation, the device will send printing information to the sewed sensors. The sewed sensor will forward the request of printing to the printing device.

Fig. 5.9 Future applications of wireless body area network (Future Applications of WBAN 2018)

5.5.2 *Use of Biomedical Sensors in Wireless Body Area Networks*

The biomedical sensor is made of two units that are radio transceiver and physiological-signal sensor (Chan et al. 2012). The function of the radio transceiver is to receive and transmit signals from sensor node and forward toward the destination node, while function of the physiological-signal sensor is to convert analog signal into digital form that has sensed from sensing environment. Further, the existing BMSs are respiration, blood pressure, accelerometer, heart rate, glucose, ECG, EEG, blood oxygen, EMG, pulse oximetry, pressure, gyroscope, and motion sensors used to monitor different vital signs of a patient. The functionality of some BMSs with their respective transfer rates (data rate) is presented in Table 5.1 in the following.

These BMSs help in monitoring vital signs and transmitting alert signals to the medical doctor if the condition of a patient is life threatening. Further, every BMS has different transfer rates in transmission of sensory data, and they need different quality of services (QoSs) due to their sensitive data. Therefore, they need to allocate

Table 5.1 Functionalities of BMSs in WBAN (Chakraborty et al. 2013; Al Ameen et al. 2012)

Name of BMS	Transfer rate	Use
Respiration	0.23–9.95 kbps	Respiration assists different organs to consume energy with the support of oxygen and glucose
Blood pressure	<9.99 bps	This sensor contains systolic and diastolic values which show either blood pressure level is normal or abnormal
Accelerometer	11 kbps	Assisting to show directions in 3D for calculating the desired energy for movements
Heartbeat	2.2 kbps	This sensor monitors beat of heart per minute showing low and high threshold values or normal ranges
ECG	142 kbps with 12 leads	Contains of different wires installed on chest to monitor heart rate
Temperature	122 bps	Monitors the patient's body to show either it has coldness or hotness
Blood oxygen	15 kbps	Needs a certain amount of oxygen for smooth flow of blood in the body
Gyroscope	9 kbps	Monitors changes in body and sends an alert signal during emergency situation
Pressure	2.2 kbps	Normally, this sensor is placed on the shoulder of a patient to monitor sitting and falling position
Pulse oximetry	1.22–2.1 kbps	Assures delivery of certain level of oxygen in blood
EMG	318–580 kbps	Uses electrodes to monitor neuromuscular and is placed in the human body
EEG	30 kbps	Brain uses waves to recognize whether it is working in normal or in abnormal conditions

dedicated paths and channels for data transmission without collision, retransmission, and delay and consume minimum energy in transmission of data with higher data reliability. However, the research community needs to develop innovative solutions to transmit data without retransmission due to collision and delay. In addition, BMSs need to consume minimum energy in different decisions of path selection.

5.6 Existing Challenges in Wireless Body Area Networks

IEEE has defined two standards for wireless communication that are IEEE 802.15.4 (IEEE 802.15.4 2006) and IEEE 802.15.6 (Man et al. 2012). IEEE publicizes guidelines on MAC and PHY layers, while there are no guidelines provided about routing layer. Hence, it is the responsibilities of the research community to design and develop routing protocols which must be delay-aware, energy-efficient, and a highly reliable in path selections. Furthermore, IEEE 802.15.4 has designed and developed for wireless sensor networks (WSNs), while the Task Group 6 (TG6) has specially designed and developed IEEE 802.15.6 for wireless body area networks (WBANs). IEEE has published its first draft related to IEEE 802.15.6 WBAN in 2012. However, the research community and academia have been used IEEE 802.15.4 for WBAN before commencing of IEEE 802.15.6. IEEE 802.15.4 has provided all functionalities which are currently provided by IEEE 802.15.6. Numerous researches have been conducted for WBAN, and various routing protocols and MAC protocols have been suggested which are explained in the following.

5.6.1 Routing Protocols

The routing protocols assist to specify devices for communication with other devices and distribution of network information which enables to select a reliable path between devices. Normally, the sensory data of patient's body is classified into non-emergency and emergency data. The non-emergency data contains normal readings of vital signs like normal temperature and blood pressure, while emergency data contains abnormal reading of vital signs such as abnormal readings of heartbeat rate and respiratory rate. This sensory data is transmitted to the coordinator without priority, and the coordinator further transmits them without prioritization of emergency data. However, the existing classification of the patient's data is not sufficient because it does not distinguish between same types of emergency data if two different sensors transmit them to the coordinator at the same time. In addition, the coordinator is not able to resolve the conflict of allocation of path/channel on the priority if two sensors transmit data at the same time.

The selection of appropriate paths is based on the residual energy level of nodes, avoidance of the hot spot, and delay paths. Due to this, the research communities divide the routing protocols in WBAN into four groups (Bangash et al. 2014), as

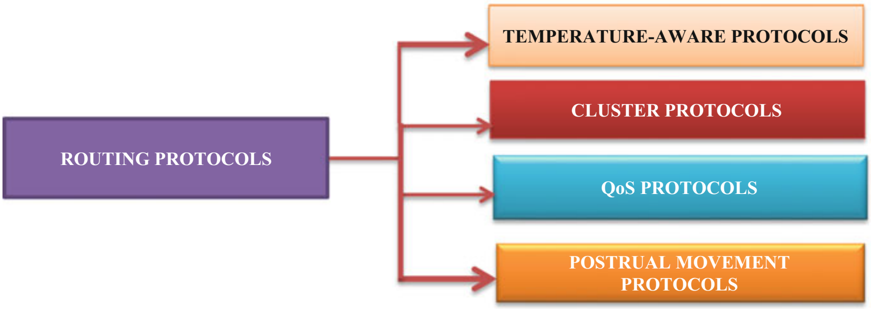

Fig. 5.10 Classification of routing protocols in WBAN

shown in Fig. 5.10. These are temperature-aware routing protocols, cluster routing protocols, QoS routing protocols, and postural movement protocols. Explanation of each routing protocol is presented in the following.

5.6.1.1 Temperature-Aware Routing Protocols

The monitoring of vital signs of the patient's body is with the support of BMSs. A BMS heats up during monitoring and transmission of sensory data of vital signs using multiple BMSs which burns skin and tissues. Causes of heat-up of BMS are using of high-frequency radio power, radiation of antenna, and the node's circuitry (Movassaghi et al. 2014). Numerous studies have suggested new design and development of routing protocols. Further, the researchers need to design and develop energy-efficient algorithms, selection of reliable paths with lowest temperature rise and shortest path to destination.

5.6.1.2 Cluster Routing Protocols

The large scale of communication area is divided into small area, known as cluster-based routing protocols. The aim of clustering is to provide optimal connectivity among nodes in WBAN. For example, the deployed BMS can communicate with other BMSs in one sitting position of person. In case of changing of sitting position of person, a BMS cannot communicate to other BMSs due to non-line-of-sight, and BMSs consume maximum energy by using higher power of antenna, which may cause other problems as mentioned in temperature-aware routing protocols.

5.6.1.3 Quality of Service (QoS)-Based Routing Protocols

The problems of QoS arises in communication network when outnumber of BMSs demand for dedicated and guaranteed bandwidth and network cannot fulfill requirements of the BMSs. In this situation, the congestion inside network has increased by dropping maximum amount of data which causes delay in retransmission of data and BMSs consumes a high amount of energy. The researchers need to design a reliable and efficient algorithms based on QoS.

5.6.1.4 Postural Movement-Based Routing Protocols

Postural means bringing change in the static object. In WBAN, the structure of the star or mesh topology has been frequently changed due to physical change in the body such as sit, lying down, stand, walk, defective sitting on the sofa, and the defective sleeping position. Due to effect of postural movement, BMSs cannot transmit data directly to the coordinator by consuming a high energy with the support of relay nodes. Existing researches have suggested to use line-of-sight (LOS), non-line-of-sight (NLOS) (Latré et al. 2010a), and store-and-forward (Quwaider and Biswas 2010) techniques. However, the patient's data needs reliable paths for data transmission without delay which can extend the network lifetime.

Another most important challenge is securing of sensory data in transmission. The asymmetric encryption techniques cannot be implemented due to resource-constraint setup of the tiny BMS. However, the existing academia needs to design a lightweight security protocols for better security and optimal utilization of the hardware sources of BMS. In concluding remarks, one should carefully design a routing protocol considering different pros and cons of the four routing protocols in WBAN.

5.7 MAC Protocols

As mentioned earlier, there are two standards used for WBAN containing IEEE 802.15.4 and IEEE 802.15.6. Each of them is explained in the following.

5.7.1 Superframe Structure of IEEE 802.15.4

There are two types of superframe structure of IEEE 802.15.4 presented. First form of superframe structure is known as non-beacon mode, and the second form of superframe structure is known as beacon-enabled mode (Ullah et al. 2010). In the non-beacon-enabled mode, the allocation of channels to BMSs is purely implemented on contention with access scheme unslotted CSMA/CA. The non-beacon mode is

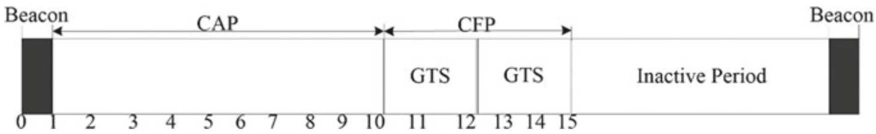

Fig. 5.11 Superframe structure of IEEE 802.15.4 (re-drawn)

specially designed for limited number of nodes and has no special consideration needed to take care of data in transmission.

The beacon-enabled-mode-based superframe of IEEE 802.15.4 (Ullah et al. 2010) is presented in Fig. 5.11 and comprises of beacon, contention-access period (CAP), contention-free period (CFP), and inactive period (IP). This superframe structure of IEEE 802.15.4 is implemented on the coordinator node in MAC layer of OSI model with sixteen channels/slots. Moreover, CAP is implemented on the scheduling access CSMA/CA scheme, while CFP is implemented on TDMA scheduling access scheme including guaranteed time slots (GTSs). The coordinator allocates CFP channels to those BMSs who got access in the CAP channels. However, the allocation of the CAP's channel is based on the contention. The IP is used for sleep period when a coordinator is free of allocation of slots.

At the beginning of superframe structure, the coordinator broadcasts a beacon frame to all BMSs in the network. This frame contains information about scanning of active and passive channels, calling of the next beacon interval, and time interval of the superframe duration. However, IEEE 802.15.4 (Touati and Tabish 2013) has the following drawbacks.

i. The superframe structure of IEEE 802.15.4 assigns sixteen channels which is not sufficient for large data produced by BMSs.
ii. Allocation of channels to BMSs is based on the contention.
iii. Allocation of the GTS CFP slots once a BMS gets access of channel in CAP.
iv. No priority has defined to allocate channel to emergency data.
v. Due to limited channels and round contention, a maximum number of collisions occur in CAP in which BMSs retransmit data with a higher delay and consume a maximum energy.
vi. Considered only emergency and non-emergency data.
vii. Gap in the research of WBAN for IEEE 802.15.4 is that it does not resolve the conflict of slot allocation if BMSs have same types of data.

Hence, the existing research community has suggested many superframe structures for MAC protocol in order to reduce collision, delay, retransmission of the lost packets, and energy consumption.

5.7.2 Superframe Structure of IEEE 802.15.6

In 2006, IEEE 802.11 had started new group that was Task Group 6 (TG6) (Al Ameen et al. 2012). TG6 was assigned a task to design low-power sensors for monitoring the health of patients. In 2012, the first draft was published for research community containing information of superframe structure of IEEE 802.15.6 MAC. Furthermore, the superframe structure of IEEE 802.15.6 comprises of beacon, exclusive access phase (EAP-I/II), random-access phase (RAP-I/II), type I/II, and contention-access period (CAP) (Rousselot and Decotignie 2009). At the beginning of communication, the coordinator transmits a beacon frame to all nodes in the network for communication convergence. EAP-I and EAP-II have designed to handle emergency traffic, while RAP-I, RAP-II, and CAP have designed to manage and handle non-emergency data of patient. Type I is associated with emergency data, while type II is associated with non-emergency data. Type indicates to coordinator which kinds of slots need to be occupied. The slotted ALOHA and CSMA/CA scheduling access schemes have implemented on IEEE 802.15.6. However, it has the same limitations as mentioned in IEEE 802.15.4 like high energy consumption, allocation of channels based on contention due to which retransmission becomes high with a higher delay, and not suitable for emergency data.

5.8 Introduction to Big Data

With the beginning of twenty-first century, the cloud computing is emerging technology for managing and allocation of resources online to different stack holders. Such stack holders are the government machinery, industry, and academia (Bates et al. 2014). Due to this importance, the big data has been introduced and merged in cloud computing to handle different input, output, and processed data of various devices for efficient utilization of resources online without buying costly equipment. Further, it has been reported in the Compliance, Governance and Oversight Council (CGOC) (Du et al. 2015) that each year volume of data becomes double and the existing infrastructure is not able to handle this huge amount of data locally. In the similar way, it has also been reported that huge amount of sensory data is produced from vital signs of a patient's body which needs efficient mechanisms to receive, process, and disseminate data toward the medical doctors. For this purpose, the existing studies in WBAN big data have numerous contributions toward an efficient design and development of mechanisms. The paper (Du et al. 2015) has included a mechanism who designed a framework in WBAN for big data for processing of sensory information of a patient, as shown in Fig. 5.12. At the beginning of first phase, a huge of amount data is collected from patient's body with the support of deployed BMSs containing clinical history from cloud servers. The load balancer unit extracts very relevant information from sensory data of patient's body about diseases based on information of clinical history. Further, distributed CEP is an emergent technology

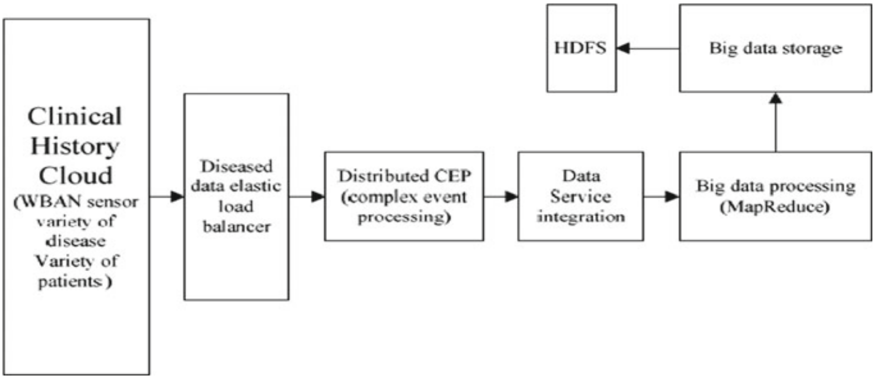

Fig. 5.12 Framework for WBAN big data (Du et al. 2015)

for WBAN in big data with functionalities to process data and identifies the actual disease(s). In this way, the data is transmitted using online data service integration. The big data processing unit removes detailed information using MapReduce feature provided by Google and finally stores data in the big data storage server. The final unit is Hadoop Distributed File System (HDFS) which splits data into different working subunits with attributes of key and values. However, the suggested WBAN big data model has many steps to extract the accurate information which creates overhead in terms of a higher delay for decision-making and consumes a maximum energy of BMSs by losing the life of patients. In addition, many WBANs have problems of frequency interference/overlapping which may not able to process the sensory data of patients and transmit to the medical doctor for optimal treatment. The authors of (Quwaider 2014) has designed conventional server (CS) and virtualized cloudlet (VC) to utilize efficiently online resources of big data for WBAN. Figure 5.13 shows collection of data using cloudlet in WBA. However, it has the same challenging issues as mentioned in (Du et al. 2015).

5.9 Applications of Big Data in WBAN

In this section, we will elaborate the applications of big data in WBAN which comprises of monitoring of vital signs and analysis, early detection of abnormal conditions of patient, and daily basis activity monitoring of a patient using BMSs (Lin et al. 2018).

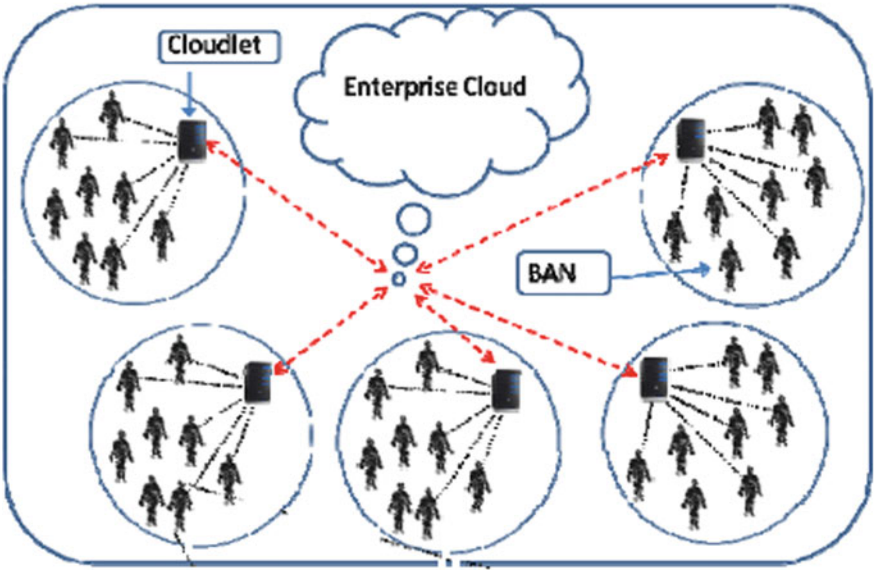

Fig. 5.13 WBAN with cloudlet-based data collection system (Quwaider 2014)

5.9.1 Monitoring of Vital Signs and Analysis

Big data plays a vital role in predication of accurate symptoms finding in patient's body using historical and clinical data information. With this predication, it improves primarily the quality of life. However, these predications are based on some certain parameters as mentioned in (Bates et al. 2014) that are management and childbirth, consumer behavior, clinical decision support, and support services. In addition, the symptoms of diseases need to be classified further for better understanding and early detection that can be performed with the support of ontology using protégé tool. Further, the existing tools which may assist to find symptoms in advance before alarming situation happens to the life of patient include Google Flu Trends and HealthMap. However, the research community needs to design and develop efficient mechanisms in early detection of abnormal conditions of patient.

5.9.2 Early Detection of Abnormal Conditions of Patient

Normally, the patient of chronic metabolic needs an adequate healthcare comparatively to other types of diseases. Due to this, different BMSs are used to monitor vital signs and periodically transmit results of sensory data to the physician. With this support, the physician can proclaim in advance the severity of patient and do precaution activities. The proclaiming practices need history and clinical data which need to be stored online and available every time. This can applicable only with the support of big data.

5.9.3 Daily Basis Activity Monitoring of a Patient Using BMSs

As mentioned earlier, Fig. 5.2 depicts how different BMSs can be deployed for monitoring of different vital signs of patient. Sometimes, physicians need to monitor the daily activities of patients and impact of these activities on their health. In addition, the physician needs sensory data of blood pressure, heart beat rate, respiratory rate, temperature, and blood glucose level.

5.10 Open Issues of WBAN

This section discusses open research issues of WBAN which comprise of routing protocols and MAC protocols.

5.10.1 Resource-Constraint Architecture of BMS

The sensor nodes have limited computational processing power, limited storage, and limited battery backup. These are challenging problems in WBAN and WSN. It has been noticed that the manufactures should need to improve the design of BMSs and include new features like harvesting for energy consumption in replacement of existing architecture.

5.10.2 Hotspot Paths

During monitoring of vital signs and data dissemination activities sometimes heat-up BMSs due to radio frequency, radiation of BMSs and circuitry design of BMSs. With this heat-up, the deployed BMSs burn skins and tissues of patient.

5.10.3 *QoS in WBAN*

The sensory data of patient's body is divided into critical, delay-sensitive, reliability-sensitive, and ordinary data. This data does not accept delay in transmission toward the coordinator and must allocate guaranteed QoS.

5.10.4 *Path Loss in WBAN*

Deployed BMSs are directly (line-of-sight) connected with coordinator for data transmission. Due to postural movement, the topology structure changes frequently and BMSs loss paths for data transmission.

5.10.5 *Data Protection in WBAN*

The sensory data of patient's body is important to secure in transmission from eavesdroppers in the resource-constraint environment of BMS.

5.10.6 *Step-Down in Energy Consumption*

Due to periodically monitoring of vital signs and transmission of results consume maximum energy of BMSs. However, the research community needs to design harvesting-based energy-efficient architecture of BMSs.

5.10.7 *Channel Access Allocation and Its Complexity*

The existing techniques for channel allocation are contention and predefined. However, this is expensive practices and the delay-sensitive data of the patient's body does not accept delay in allocation of channels and transmission.

5.10.8 *Permission- and Preemption-Based Channel Assignment*

The protocols should be efficiently designed and developed to allocate channels based on permission and preemption according to the sensitivity of patient's data.

5.11 Open Issues of Big Data

This section discusses the vital research issues of big data, which are data management which is elaborated in the following.

5.11.1 Varieties of Data

On daily basis, every IoT-enabled device generates millions of data. Now, how to extract the important and the most relevant information from unstructured of off-line and streaming data quickly?

5.11.2 Increased Amount of Data Storage

How to design and develop efficient methods to recognize the important and relevant data from large amount of unstructured data?

5.11.3 Integration of Data from Different Sources

The research community needs to design new protocols to combine relevant data from same objects and integrate them for further use.

5.11.4 Allocation of Channel, Processing, and Management of Data

Need to improve the design of the existing communication models for off-line and online streaming of data for efficient allocation of resources, process, and manage them for accurate retrieval of information on time.

5.11.5 Cost-Effective Business Model

The future models for businesses should be user-friendly for available resources with minimum cost.

5.11.6 Delay-Aware Models for Quick Solutions

The research community should design efficient delay-aware models in terms of quick responses to customers on time.

5.11.7 Automation in Allocation of Services

The new business models should replace services provided by human with machine learning and big data analysts.

References

Acampora G, Cook DJ, Rashidi P, Vasilakos AV (2013) A survey on ambient intelligence in health care. Proc IEEE (Institute of Electrical and Electronics Engineers) 101(12):2470–2494. https://doi.org/10.1109/JPROC.2013.2262913

Al Ameen M, Ullah N, Chowdhury MS, Islam SMR, Kwak K (2012) A power efficient MAC protocol for wireless body area networks. EURASIP J Wirel Commun Netw 2012(1):33. https://doi.org/10.1186/1687-1499-2012-33

Applications of WBAN (2018) Retrieved from https://www.slideshare.net/kunalgoel1984/wsn-in-body-area-network-based-on-egergy-conservation

Bangash JI, Abdullah AH, Anisi MH, Khan AW (2014) A survey of routing protocols in wireless body sensor networks. Sensors (Basel, Switzerland) 14(1):1322–1357. https://doi.org/10.3390/s140101322

Bates DW, Saria S, Ohno-machado L, Shah A (2014) Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff 33(7):1–10. https://doi.org/10.1377/hlthaff.2014.0041

Cavallari R, Martelli F, Rosini R, Buratti C, Verdone R (2014) A survey on wireless body area networks: technologies and design challenges. IEEE Commun Surv Tutorials 16(3):1–23. https://doi.org/10.1109/SURV.2014.012214.00007

Chakraborty C, Gupta B, Ghosh SK (2013) A review on telemedicine-based WBAN framework for patient monitoring. Telemed E-Health 19(8):619–626. https://doi.org/10.1089/tmj.2012.0215

Chan M, Estève D, Fourniols J-Y, Escriba C, Campo E (2012) Smart wearable systems: current status and future challenges. Artif Intell Med 56(3):137–156. https://doi.org/10.1016/j.artmed.2012.09.003

Di Cataldo S, Bottino A, Islam IU, Vieira TF, Ficarra E (2014) Subclass discriminant analysis of morphological and textural features for hep-2 staining pattern classification. Pattern Recogn 47(7):2389–2399. https://doi.org/10.1016/j.patcog.2013.09.024

Du Y, Hu F, Wang L, Wang F (2015) Framework and challenges for wireless body area networks based on big data. In: IEEE 2015 IEEE international conference on digital signal processing (DSP), pp 497–501 https://doi.org/10.1109/ICDSP.2015.7251922

El Hihi S, Bengio Y (1996) Hierarchical recurrent neural networks for long-term dependencies. In: Advances in neural information processing systems, pp 493–499

Future Applications of WBAN (2018). Retrieved from https://www.renesas.com/us/en/about/edge-magazine/solution/07-wireless-sensor-networks.html

Hadsell R, Sermanet P, Ben J, Erkan A, Scoffier M, Kavukcuoglu K et al. (2014) Learning long-range vision for autonomous off-road driving, 2009 26(2):120–144

He Y, Zhu W, Guan L (2011) Optimal resource allocation for pervasive health monitoring systems with body sensor networks. IEEE Trans Mob Comput 10(11):1558–1575. https://doi.org/10.1109/TMC.2011.83

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778

Hochreiter S, Schmidhuber J (1997). Long short-term memory. Neural Comput 8(9):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

IEEE 802.15.4 (2006) IEEE standard for Information Technology 802.15.4, wireless medium access control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (LR-WPANs). Institute of Electrical and Electronics Engineers, New York, US, vol 2006. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Part+15.4:+Wireless+Medium+Access+Control+(MAC)+and+Physical+Layer+(PHY)+Specifications+for+Low-Rate+Wireless+Personal+Area+Networks+(WPANs)#0%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Sear

Karen Simonyan AZ (2014) Very deep convolutional networks for large-scale image recognition. Comput Vision Patt Recogn. Retrieved from https://arxiv.org/pdf/1409.1556.pdf

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105

Latré B, Braem B, Moerman I, Blondia C, Demeester P (2010a) A survey on wireless body area networks. Wirel Netw 17(1):1–18. https://doi.org/10.1007/s11276-010-0252-4

Latré B, Braem B, Moerman I, Blondia C, Demeester P (2010b, November 11) A survey on wireless body area networks. Wirel Netw. https://doi.org/10.1007/s11276-010-0252-4

Lecun Y, Bottou L, Bengio Y, Haffner P (n.d.) Gradient-based learning applied to document recognition. In: Proceedings of the IEEE, pp 2278–2324. IEEE. https://doi.org/10.1109/5.726791

Lin R, Ye Z, Wang H, Wu B, Technology S (2018) Chronic diseases and health monitoring big data: a survey. IEEE Rev Biomed Eng 3333(c):1–15. https://doi.org/10.1109/RBME.2018.2829704

Maas AL, Hannun AY (2013) Rectifier nonlinearities improve neural network acoustic models. Proc icml 1–20

Man LAN, Committee S, Computer I (2012) IEEE standard for local and metropolitan area networks: Part 15. 6: Wireless Body Area Networks IEEE Computer Society

Movassaghi S, Member S, Abolhasan M, Member S (2014) Wireless body area networks: a survey. IEEE Commun Surv Tutorials 16(3):1658–1686

Multi-column deep neural network for traffic sign classification (2012) Neural Netw 32:333–338

Murray CJL, Vos T, Lozano et al (2012) Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380(9859):2197–223. https://doi.org/10.1016/S0140-6736(12)61689-4

Quwaider M (2014) An efficient big data collection in body area networks. In: 2014 5th international conference on information and communication systems (ICICS), pp 1–7. Irbid, Jordan. https://doi.org/10.1109/IACS.2014.6841986

Quwaider M, Biswas S (2010) DTN routing in body sensor networks with dynamic postural partitioning. Ad Hoc Netw 8(8):824–841. https://doi.org/10.1016/j.adhoc.2010.03.002

Quwaider M, Jararweh Y (2015) A cloud supported model for efficient community health awareness. Pervasive Mob Comput 28:35–50. https://doi.org/10.1016/j.pmcj.2015.07.012

Rousselot J, Decotignie J-D (2009) Wireless communication systems for continuous multiparameter health monitoring. In: 2009 IEEE international conference on ultra-wideband, 2009, pp 480–484. https://doi.org/10.1109/ICUWB.2009.5288747

Sainath TN, Mohamed AR, Kingsbury B, Ramabhadran B (2013) Deep convolutional neural networks for LVCSR. In: 2013 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 8614–8618

Shu M, Yuan D, Zhang C, Wang Y, Chen C (2015) A MAC protocol for medical monitoring applications of wireless body area networks. Sensors 15(6):12906–12931. https://doi.org/10.3390/s150612906

Sutskever I, Martens J, Hinton GE (2011). Imagenet: a large-scale hierarchical image database. In: Proceedings of the 28th international conference on machine learning (ICML-11), pp 1017–1024

Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826

Taigman Y, Yang M, Ranzato MA, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1701–1708

Touati F, Tabish R (2013) u-Healthcare system: state-of-the-art review and challenges. J Med Syst 37(3):9949. https://doi.org/10.1007/s10916-013-9949-0

Ul-Islam I (2014) Feature fusion for pattern recognition. Politecnico di Torino. Dipartimento di automatica e informatica. Retrieved from https://books.google.com.pk/books/about/Feature_Fusion_for_Pattern_Recognition.html?id=GT08ngAACAAJ&redir_esc=y

Ullah S, Shen B, Islam SMR, Khan P, Saleem S, Kwak KS (2010) A Study of medium access control protocols for wireless body area networks, pp 1–13. Retrieved from http://arxiv.org/abs/1004.3890v2

Index

M. Khan et al., *Deep Learning: Convergence to Big Data Analytics*, SpringerBriefs in Computer Science, https://doi.org/10.1007/978-981-13-3459-7