Industry 4.0 Interoperability, Analytics, Security, and Case Studies
Industry 4.0 or fourth industrial revolution refers to interconnectivity, automation and real time data exchange between machines and processes. There is a tremendous growth in big data from internet of things (IoT) and information services which drives the industry to develop new models and distributed tools to handle big data. Cutting-edge digital technologies are being harnessed to optimize and automate production including upstream supply-chain processes, warehouse management systems, automated guided vehicles, drones etc. The ultimate goal of industry 4.0 is to drive manufacturing or services in a progressive way to be faster, effective and efficient that can only be achieved by embedding modern day technology in machines, components, and parts that will transmit real-time data to networked IT systems. These, in turn, apply advanced soft computing paradigms such as machine learning algorithms to run the process automatically without any manual operations. The new book series will provide readers with an overview of the state-of-the-art in the field of Industry 4.0 and related research advancements. The respective books will identify and discuss new dimensions of both risk factors and success factors, along with performance metrics that can be employed in future research work. The series will also discuss a number of real-time issues, problems and applications with corresponding solutions and suggestions. Sharing new theoretical findings, tools and techniques for Industry 4.0, and covering both theoretical and application-oriented approaches. The book series will offer a valuable asset for newcomers to the field and practicing professionals alike. The focus is to collate the recent advances in the field, so that undergraduate and postgraduate students, researchers, academicians, and industry people can easily understand the implications and applications of the field.

Industry 4.0 Interoperability, Analytics, Security, and Case Studies

G Rajesh, Raajini X. Mercilin and Hein Dang Thi Thu

For more information about this series, please visit: https://www.routledge.com/Big-Data-for-Industry-4.0-Challenges-and-Applications/book-series/CRCBDICA
Industry 4.0 Interoperability, Analytics, Security, and Case Studies

Edited by
G. Rajesh, X. Mercilin Raajini, and Hien Dang
Contents

Editors .................................................................................................................................................. vii
Contributors ......................................................................................................................................... ix

Chapter 1  Big Data Analytics and Machine Learning for Industry 4.0:
An Overview .......................................................................................................................................... 1

*Nguyen Tuan Thanh Le and Manh Linh Pham*

Chapter 2  Impact of Blockchain-Based Cyber Security Implementing
Industry 4.0 ............................................................................................................................................. 13

*Puja Das, K. Martin Sagayam, Asik Rahaman Jamader, and Hung Nguyen*

Chapter 3  Hybrid Huffman Tree Model for Securing Medical
Data Using Steganography ..................................................................................................................... 35

*A. Saranya, Mohak Narang, R. Anandan, S. Aruna, and Balaganesh*

Chapter 4  Energy Conservation in IOT-Based Intelligent
Transport System ..................................................................................................................................... 51

*A. Sivasangari, D. Deepa, Chiung Ching Peter Ho, M.S. Roobini, and R. Vignesh*

Chapter 5  Digital Information System for Attack Reduction
in Industry 4.0 ........................................................................................................................................ 65

*Srinivasan Selvaraj, Raja Kothandaraman, Kannan Kaliyan, and Pratheepan Thuraiyappah*

Chapter 6  An Intelligent Blockchain-Based Framework
for Securing Smart Grid ......................................................................................................................... 81

*G. Jaspher W. Kathrine, Bahman Javadi, J. Dinesh Peter, and G. Matthew Palmer*

Chapter 7  Intelligent Fog Computing for Industrial
Wireless Sensor Networks ..................................................................................................................... 95

*S. Karthikeyan, K. Vimala Devi, and K. Valarmathi*
Chapter 8  A Secure Data Sharing Scheme Based on Blockchain for Industrial Internet of Things Using Consensus Algorithm

M. Yuvaraju and P.M. Benson Mansingh

Chapter 9  Smart Trust Management Scheme for Detecting Attacks in IoT and Ubiquitous Computing

A. Sivasangari, R.M. Gomathi, A. Jesudoss, L. Lakshmanan, K. Indira, and T. Samraj Lawrence

Chapter 10  IoT-Based Smart Pipeline Leakage Detecting System for Petroleum Industries

A. Devi, M. Julie Therese, P. Dharani Devi, and T. Ananth Kumar

Chapter 11  IOT-Based Smart Irrigation Systems

M.R. Ebenezar Jebarani, P. Kavipriya, S. Lakshmi, and Nehru Kandasamy

Chapter 12  Safety Wing for Industry (SWI 2020) – An Advanced Unmanned Aerial Vehicle Design for Safety and Security Facility Management in Industries

T. Ananth Kumar, S. Arunmozhi Selvi, R.S. Rajesh, and G. Glorindal

Chapter 13  An Efficient Decentralized Medical Prescription Tracking Using Blockchain


Chapter 14  Blockchain for Industry 4.0: An Assessment of Blockchain Adoptability

Rajakumar Arul, Raja Kothandaraman, Kannan Kaliyan, and Nizar Banu P.K.

Chapter 15  An Effective E-Learning Mechanism to Meet the Learning Demand of Industry 4.0

A. Jaya, P. Sheik Abdul Khader, A. Abdul Azeez Khan, K. Javubar Satthick, L. Arun Raj, and Ho Chiung Ching

Index
Editors

**G. Rajesh** is an Assistant Professor in the Department of Information Technology at Anna University, Chennai, India. He earned a PhD in trust-based temporal data aggregation for energy-constrained wireless sensor networks at Anna University in 2016. He did postgraduate in computer science engineering at the Government College of Engineering, Tirunelveli under Anna University. He earned an undergraduate degree in information technology at AVC College of Engineering at Anna University. He has 12 years of teaching and research experience. His area of research interest includes wireless sensor networks and its IoT applications, software engineering, machine learning, data analysis, and computational optimization. His research contributions include 3 patents from India, 2 authored books, 2 edited books, and many articles in reputed international journals and at conferences. He also is a reviewer at reputed journals and conferences. He is an active member of IEEE, SAEINDIA, the Society of Automotive Engineers, and the International Association of Engineers (IAENG).

**X. Mercilin Raajini** is an Associate Professor in the Department of Electronics and Communication at Prince Shri Venkatesahwara Padmavathy Engineering College at Anna University, Chennai, India. She earned a PhD in enhanced routing techniques for energy management in wireless sensor networks at Anna University, India in 2016. She completed postgraduate work in power electronics and drives at Government College of Engineering, Tirunelveli at Anna University in 2007. She earned an undergraduate degree in electronics and communication at Noorul Islam College of Engineering, under MS University in 2001. She has 15 years of teaching and research experience. Her area of research interest includes wireless sensor networks and its IoT applications, signal processing, machine learning, and computational optimization. Her research contributions include 2 patents from India, 3 authored books, 2 edited books, and many published articles in international journals and at conferences. She also is a reviewer for reputed journals and conferences. She is an active member of ISTE, SAEINDIA, and the Society of Automotive Engineers.
Hien Dang earned a PhD in computer science in 2010. She is on the faculty at Computer Science and Engineering Faculty, Thuy Loi University, Vietnam. She is also a research scholar at the University of Massachusetts, Boston, USA. She is the author or coauthor of many books and has more than thirty papers in journals and conference contributions. Her research areas include artificial intelligence, machine learning, deep learning, problems about regressive prediction, Big Data analytics, and processing for practical applications.
Contributors

Anandan
Department of Computer Science & Engineering
Vels Institute of Science, Technology & Advanced Studies
Chennai, India

Puja Das
Department of Computer Science
Hiralal Majumdar Memorial College for Women
Kolkata, India

Rajakumar Arul
Amrita School of Engineering
Amrita Vishwa Vidyapeetham
Bangalore, India

D. Deepa
School of Computing
Sathyabama Institute of Science and Technology
Chennai, India

S. Aruna
Department of Software Engineering
SRM Institute of Science & Technology
India

A. Devi
Department of ECE
IFET College of Engineering
Villupuram, India

Balaganesh
Faculty of Computing Science and Multimedia
Lincoln University College
Malaysia

K. Vimala Devi
School of Computer Science and Engineering
Vellore Institute of Technology
India

P. K. Nizar Banu
Department of CS
Christ Deemed to Be University
Bangalore, India

P. Dharani Devi
Department of ECE
IFET College of Engineering
Villupuram, India

Ho Chiung Ching
Department of Computing and Information Systems
Sunway University
Malaysia

G. Glorindal
Department of Information and Communication Technology
DMI – St. John the Baptist University
Malawi

Peter Ho Chiung Ching
Department of Computing and Information Systems
School of Science and Technology
Sunway University
Selangor, Malaysia

R.M. Gomathi
School of Computing
Sathyabama Institute of Science and Technology
Chennai, India
Contributors

K. Indira
School of Computing
Sathyabama Institute of Science and Technology
Chennai, India

Nehru Kandasamy
Department of Electrical & Computer Engineering
National University of Singapore
Singapore

Asik Rahaman Jamader
Penguin School of Hotel Management
Kolkata, India

S. Karthikeyan
Department of Electronics and Instrumentation Engineering
M. Kumarasamy College of Engineering
Karur, India

Bahman Javadi
School of Computer, Data and Mathematical Sciences
Western Sydney University
Australia

G. Jaspher W. Kathrine
Department of CSE
Karunya Institute of Technology and Sciences
Coimbatore, India

A. Jaya
Department of Computer Applications
B.S. Abdur Rahman Crescent Institute of Science & Technology
Chennai, India

P. Kavipriya
Department of Electronics and Communication Engineering
Sathyabama Institute of Science and Technology
Chennai, India

M.R. Ebenezar Jebarani
Department of Electronics and Communication Engineering
Sathyabama Institute of Science and Technology
Chennai, India

P. Sheik Abdul Khader
Department of Computer Applications
B. S. Abdur Rahman Crescent Institute of Science & Technology
Chennai, India

A. Jesudoss
School of Computing
Sathyabama Institute of Science and Technology
Chennai, India

A. Abdul Azeez Khan
Department of Computer Applications
B. S. Abdur Rahman Crescent Institute of Science & Technology
Chennai, India

Kannan Kaliyan
Faculty of Engineering and Technology
Jain Deemed to Be University
Bangalore, India
Raja Kothandaraman  
Department of Computer Science and Engineering  
SRM Institute of Science and Technology  
India

T. Ananth Kumar  
Department of ECE  
IFET College of Engineering  
Villupuram, India

L. Lakshmanan  
School of Computing  
Sathyabama Institute of Science and Technology  
Chennai, India

S. Lakshmi  
Department of Electronics and Communication Engineering  
Sathyabama Institute of Science and Technology  
Chennai, India

T. Samraj Lawrence  
Department of IT  
Dambi Dolo University  
Ethiopia

Nguyen Tuan Thanh Le  
Faculty of Computer Science and Engineering,  
Thuyloi University  
Hanoi, Vietnam

D.R. Anita Sofia Liz  
Department of Computer Science and Engineering  
New Prince Shri Bhavani College of Engineering and Technology  
Chennai, India

P.M. Benson Mansingh  
Department of Electronics and Communication Engineering  
Sri Ramakrishna Institute of Technology  
Coimbatore, India

Mohak Narang  
Department of Software Engineering  
SRM Institute of Science & Technology  
Chennai, India

Hung Nguyen  
Model Validation Analyst  
Regions Financial Corp.  
Birmingham, Alabama

G. Matthew Palmer  
Department of CSE  
Karunya Institute of Technology and Sciences  
Coimbatore, India

J. Dinesh Peter  
Department of CSE  
Karunya Institute of Technology and Sciences  
Coimbatore, India

Manh Linh Pham  
Faculty of Information Technology  
VNU University of Engineering and Technology  
Hanoi, Vietnam

L. Arun Raj  
Department of Computer Applications  
B. S. Abdur Rahman Crescent Institute of Science & Technology  
Chennai, India
Contributors

R.S. Rajesh
Department of Computer Science and Engineering
Manonmaniam Sundaranar University
Tirunelveli, India

K. Javubar Sathick
Department of Computer Applications
B. S. Abdur Rahman Crescent Institute of Science & Technology
Chennai, India

A. Kapil Ramachandran
Department of Computer Science and Engineering
New Prince Shri Bhavani College of Engineering and Technology
Chennai, India

S. Arunmozhi Selvi
Department of Computer Science and Engineering
Manonmaniam Sundaranar University
Tirunelveli, India

M.S. Roobini
School of Computing
Sathyabama Institute of Science and Technology
Chennai, India

Srinivasan Selvaraj
School of Computer Science & IT
Jain Deemed to Be University
Bangalore, India

M.S. Roobini
School of Computing
Sathyabama Institute of Science and Technology
Chennai, India

A. Sivasangari
School of Computing
Sathyabama Institute of Science and Technology
Chennai, India

H. Saagarika
Department of Computer Science and Engineering
New Prince Shri Bhavani College of Engineering and Technology
Chennai, India

M. Julie Therese
Department of ECE
Sri Manakula Vinayagar Engineering College
Pondicherry, India

K. Martin Sagayam
Department of ECE
Karunya Institute of Technology and Sciences
Coimbatore, India

Pratheepan Thuraiyappah
Library Department
Uva Wellassa University
Badulla
Sri Lanka

A. Saranya
Department of Computer Science & Engineering
Vels Institute of Science, Technology & Advanced Studies
and
Department of Software Engineering
SRM Institute of Science & Technology
Chennai, India

K. Valarmathi
Department of Electronics and Communication Engineering
P.S.R. Engineering College
Sivakasi, India
Contributors

R. Vignesh
School of Computing
Sathyabama Institute of
Science and Technology
Chennai, India

M. Yuvaraju
Department of Electrical and
Electronics Engineering
Anna University Regional
Campus
Coimbatore, India

M. Stefi Vinciya
Department of Computer Science and
Engineering
New Prince Shri Bhavani College of
Engineering and Technology
Chennai, India
1 Big Data Analytics and Machine Learning for Industry 4.0: An Overview

Nguyen Tuan Thanh Le and Manh Linh Pham

CONTENTS

1.1 Big Data Analytics for Industry 4.0 ................................................................. 1
  1.1.1 Characteristics of Big Data ........................................................................ 1
  1.1.2 Characteristics of Big Data Analytics ...................................................... 3
1.2 Machine Learning for Industry 4.0 ................................................................. 4
  1.2.1 Supervised Learning .............................................................................. 5
  1.2.2 Unsupervised Learning ......................................................................... 5
  1.2.3 Semi-Supervised Learning ..................................................................... 6
  1.2.4 Reinforcement Learning ........................................................................ 6
  1.2.5 Machine Learning for Big Data .............................................................. 6
1.3 Deep Learning for Industry 4.0: State of the Art ........................................... 7
1.4 Conclusion ....................................................................................................... 9
Acknowledgments ................................................................................................ 9
References ............................................................................................................ 9

1.1 BIG DATA ANALYTICS FOR INDUSTRY 4.0

1.1.1 CHARACTERISTICS OF BIG DATA

The concept of “Big data” was mentioned for the first time by Roger Mougalas in 2005 [1]. It refers to a large scale data, one of the characteristics of Industry 4.0, that cannot be stored in a single computer and is almost impossible to be handled using traditional data analytics approaches. Big data applications exploded after 2011 are related to the improvement in computing power and storage as well as the reduction in the cost of sensors, communication and, recently, the development of the Internet of Things (IoT). These advances have led to the utilization of multiple sources (sensors, applications, people, and animals) in the generation of data. In 2011, Big data is defined by [2] using 4Vs characteristics, including: Volume, Velocity, Variety, and also Value. Then the fifth one, Veracity, was introduced in 2012 [3], as shown in Fig. 1.1.
FIGURE 1.1 5Vs Characteristics of Big Data

*Volume* hints to the size and/or scale of datasets. Until now, there is no universal threshold for data volume to be considered as big data, because of the time and diversity of datasets. Generally, big data can have the volume starting from exabyte (EB) or zettabyte (ZB) [4].

*Variety* implies the diversity of data in different forms which contain *structured*, *semi-structured*, or *unstructured* ones. Real-world datasets, coming from heterogeneous sources, are mostly under unstructured or semi-structured form that makes the analysis challenging because of the inconsistency, incompleteness, and noise. Therefore, data prepossessing is needed to remove noise, which includes some steps such as *data cleaning*, *data integrating*, and *data transforming* [5].

*Velocity* indicates the speed of processing data. It can fall into three categories: streaming processing, real-time processing, or batch processing. This characteristic emphasizes that the speed of producing data should keep up with the speed of processing data [4].

*Value* alludes to the usefulness of data for decision making. Giant companies (e.g., Amazon, Google, Facebook, etc.) analyze daily large scale datasets of users and their behavior to give recommendations, improve location services, or provide targeted advertising, etc. [3].

*Veracity* denotes the quality and trustworthiness of datasets. Due to the variety characteristic of data, the accuracy and trust become harder to accomplish and they play an essential role in applications of big data analytics (BDA). As with analyzing millions of health care entries in order to respond to an outbreak that impacts on a huge number of people (e.g., the CoVid-19 pandemic) or veterinary records to guess the plague in swine herd (e.g., African swine fever or porcine reproductive and
respiratory syndrome), any ambiguities or inconsistencies in datasets can impede the precision of analytic process [3], leading to a catastrophic situation.

Generally, big data in the context of Industry 4.0 can originate from several and various sources, such as: product or machine design data, machine-operation data from control systems, manual-operation records performed by staff, product-quality and process-quality data, manufacturing execution systems, system-monitoring and fault-detection deployments, information on operational costs and manufacturing, logistics information from partners, information from customers on product utilization, feedback, and so on and so forth [6]. Some of these datasets are semi-structured (e.g., manual-operation records), a few are structured (e.g., sensor signals), and others are completely unstructured (e.g., images). Therefore, an enterprise 4.0 requires cutting-edge technologies that can fully take advantage of the valuable manufacturing data, including: machine learning (ML) and BDA.

1.1.2 Characteristics of Big Data Analytics

BDA can be referred to as “the process of analyzing large scale datasets in order to find unknown correlations, hidden patterns, and other valuable information which is not able to be analysed using conventional data analytics” [7], as the conventional data analysis techniques are no longer effective because of the special characteristics of big data: massive, heterogeneous, high dimensional, complex, erroneous, unstructured, noisy, and incomplete [8].

BDA has attracted attention from not only academic but also industrial scientists as the requirement of discovering hidden trends in large scale datasets increases. Reference [9] compared the impact of BDA for Industry 4.0 with the invention of the microscope and telescope for biology and astronomy, respectively. Recently, the considerable development in the ubiquitous IoT (i.e., Internet of Things), sensor networks, and CPS (i.e., cyber-physical systems) have expanded the data-collection process to an enormous scale in numerous domains, including: social media, smart cities, education, health care, finance, agriculture, etc. [3].

Various advanced techniques to analyze data (i.e., ML, computational intelligence, data mining, natural language processing) and potential strategies (i.e., parallelization, divide and conquer, granular computing, incremental learning, instance selection, feature selection, and sampling) can help to handle big data issues. Empowering more efficient processing, and making better decisions can also be obtained by using these techniques and strategies [3].

Divide and conquer helps to reduce the complexity of computing problems. It is composed of three phases: firstly, it reduces the large-complex problem into several smaller, easier ones; secondly, it tries to solve each smaller problem; and finally, it combines the solutions of all the smaller problems to solve the original problem [3].

Parallelization allows one to improve computation time by dividing big problems into smaller instances, distributing smaller tasks across multiple threads and then performing them simultaneously. This strategy decreases computation time instead of total amount of work because multiple tasks can be performed simultaneously rather than sequentially [10].
**Incremental learning** is widely practiced and used to handle streaming data. It is a learning algorithm and can be trained continuously with additional data rather than current ones. In the learning process, this strategy tunes parameters each time new input data comes in [10].

**Granular computing** helps to simplify the elements from a large space by grouping them into subsets, or granules [11, 12]. By reducing large elements to a search space which is smaller, any uncertainty of elements in this search space is identified effectively [13].

**Feature selection** is useful for preparing high scale datasets. This strategy handles big data by determining a subset of relevant features which are for an aggregation. Nevertheless, the data representation is more precise in this particular strategy [14].

**Instance selection** is a major approach for pre-processing data. It helps to shorten training sets and run-time in the training phases [15].

**Sampling** is a method for data reducing that helps to derive patterns in big datasets by generating, manipulating, and analyzing subsets of the original data [10].

### 1.2 MACHINE LEARNING FOR INDUSTRY 4.0

Machine learning (ML), a state-of-the-art subfield of Artificial Intelligence (AI) – Fig. 1.4, has now been powering several aspects of our society: information search on the Internet, content filtered on social networks, recommendations of e-commercial platforms, or accurate language translation, virtual classrooms in education, support for diagnosing diseases in medicine, etc. [16]. It has been applied successfully to solve several real problems, such as: transcribing speech into text, matching new items, identifying objects, selecting relevant search results, etc. [16].

Actually, the goal of a typical ML to find a *mathematical formula* (i.e., the model), when applied to a collection of inputs (i.e., the training data) then produces the desired outputs [17]. The “invented” mathematical formula is also expected to generate the “correct” outputs for most other new inputs (distinct from the training data) on the assumption that those inputs come from the same or a similar statistical distribution of the training data [17].

In order to teach the machine, three components are needed, including:

- (1) **data** – the more diverse and bigger the data, the better the result; (2) **features** – also know as parameters or variables (e.g., age, gender, stock price, etc.), they are the factors that the machine is looking at; and (3) **algorithms** – the steps we follow to solve the given problem that affects the precision, performance, and size of our model [18].

Generally, ML algorithms can be classified into four main types: (1) **supervised learning**, (2) **unsupervised learning**, (3) **semi-supervised learning**, and (4) **reinforcement learning** [17, 19], as shown in Fig. 1.2.
1.2.1 SUPERVISED LEARNING

To find out a mathematical formula that maps inputs to already-known outputs, when provided with a set of human-annotated examples, is the purpose of a supervised learning algorithm. In this case, we have a “supervisor” or a “teacher” who gives the machine all the answers, like whether it’s a cat or a dog in a given picture [18], i.e., a classification problem. The “teacher” has already labeled input datasets and the machine will learn on top of these examples [18].

A specific kind of supervised learning is self-supervised learning, in which the machine learns without human-annotated labels [19]. There are still associated labels, but they are generated from the input data typically using a heuristic algorithm [19].

1.2.2 UNSUPERVISED LEARNING

In contrast to the former, in this case, the machine has no “supervisor” or “teacher”. Input data is not labeled; the machine is left on its own, trying to find certain hidden patterns or structures in datasets. For example, in a clustering problem, the model (i.e., a mathematical formula) will output a cluster identifier for each input data. Or in a dimensionality reduction problem, the model will output a new vector with fewer features than the original one of input data [17].
1.2.3 **Semi-Supervised Learning**

In this case, input data contains not only labeled but also unlabeled examples. The purpose of semi-supervised learning is the same as the purpose of supervised learning [17]. By using several unlabeled examples (i.e., adding more information about the problem that then reflects better the probability distribution of data), we can help our learning algorithm to find out better models [17].

1.2.4 **Reinforcement Learning**

In this case, the machine, also called an *agent*, is embedded in an environment. It is capable of observing *states* of that environment as a vector of features, and then can perform *actions* in response to these states [17]. Different actions can give different *rewards* to the agent and could also move it to another state of the environment, and so on. In contrast to supervised learning and unsupervised learning, where we operate with *static* datasets, in reinforcement learning, we work with *dynamic* datasets collected repeatedly from a dynamic environment.

Learning an *optimal policy* is the target of a reinforcement learning algorithm, i.e., a function that inputs the feature vector of a state and then outputs an optimal sequence of actions to execute in that state [17], in order to maximize long-term accumulated reward. For instance, a sequence of scaling actions such as adding or removing virtual machines/containers to keep up with fluctuation of a resource’s demand for big data analytic application can be a result drawn from a reinforcement learning algorithm. The feature vector here consists of information from the application itself (e.g., current computing power, workload) and the surrounding environment (e.g., type of media, other co-located applications).

1.2.5 **Machine Learning for Big Data**

Conventional ML approaches cannot handle efficiently big data problems because of its 5Vs characteristics (i.e., high speeds, diverse sources, low-value density, large volumes, and incompleteness) [3]. Therefore, several advanced ML techniques for BDA are proposed: *feature learning*, *transfer learning*, *distributed learning*, *active learning*, and *deep learning* [3].

*Feature learning* empowers a system to figure out automatically the representations required for feature detection or classification from raw datasets [3].

*Transfer learning* allows the machine to employ knowledge which has been learned from one context to new contexts. By transferring useful information from similar domains, it efficiently improves a learner from one specific domain [20].

*Distributed learning* aims to alleviate the scalability issue of conventional ML by distributing computations on datasets among a couple of machines for scaling up the process of learning [21]. One of the platforms using this distributed approach to resolve scaling problems for multi-cloud applications was proposed by [22].

*Active learning* aims to employ adaptive data collection. In this process, parameters are adjusted automatically to gather as quickly as possible useful data for accelerating ML tasks and overpowering the problem of labeling [23].
Deep learning (DL) can be employed to extract complex and high-level abstractions of data representations. It is done by using a hierarchical, layered architecture of learning, where more abstract features (i.e., higher-level) are stated, described, and implemented on top of less abstract ones (i.e., lower-level) [24] – see Fig. 1.3(a). DL techniques can analyze and learn from an enormous amount of unsupervised data, which is suitable for BDA in which raw data is mostly unlabeled as well as uncategorized [24]. We will focus on DL for Industry 4.0 in the next section.

1.3 DEEP LEARNING FOR INDUSTRY 4.0: STATE OF THE ART

Deep learning, the most exciting branch of ML – Fig. 1.4, has been expanded on the base of classic Artificial Neural Networks (NNs). It supports computational models that, in contrast to shallow NN-like models with only a few layers, consist of multiple processing (non-linear) layers. Each layer will take charge of a different level of abstraction that helps to learn hierarchical representation of data. The functionality of DL is emulated from the operation of the neuron network in the human brain for processing ambient signals [25], with the notions of axon, synapse, dendrite – see Fig. 1.3(b). DL, with different types (e.g., Recurrent Neural Networks, Autoencoders, Convolutional Neural Networks, Deep Belief Net, etc.), has outperformed others conventional ML techniques as well as improved dramatically cutting-edge real problems in the recognition of objects, speech recognition, object detection, language translation and several other areas such as self-driving cars, genomics, games, or drug discovery, etc. [26, 16].

DL helps to discover convoluted structures in a large scale dataset by using an optimization algorithm, called backpropagation, meaning “error backward propagation”. It specifies how a model changes its (up to billions of) internal parameters. In each layer, these parameters are used to compute the representation based on the last one [16]. Most contemporary DL algorithms are based on Stochastic Gradient Descent (SGD) [27].
In addition, DL requires fairly little manual engineering, as it can profit from increases in the amount of available data and computation [16], and thus is suitable for BDA. In consequence, several hidden features, which might not be seen obviously by a human, can be exposed by using a DL model [25].

For the industry sector, in order to accelerate technologies toward smart manufacturing, equipping intelligent and high-precision systems is very important because they directly affect the efficiency of related products, reinforce productivity, and also reduce operation costs as well as maintenance expenses [25]. In this context, a DL model can play an essential role. Indeed, a wide range of applications for industry such as controlling robots, object detection and tracking, visual inspection of product lines, fault diagnosis, etc., can benefit by applying a DL model [25].

Luckow et al. [28] investigated visual inspection of product lines using Convolutional Neural Network architectures including AlexNet and GoogLeNet over different DL platforms, such as Tensorflow, Caffe, and Torch. In this work, several vehicle images, along with their annotation, in the assembly line are submitted to a DL system. Consequently, by using the Tensorflow platform, they achieved the best performance with accuracy of 94%.

Lee et al. [29] tackled the detection of faults found in the process of transferring geometric shapes on a mask to the surface of a silicon wafer and classification problem in noisy settings by employing Stacked Denoising Auto-Encoders (SdA). They help to lower the noise contained in descriptive sensory data, derived from electrically mechanical disturbances as well as carry out classifications of fault. The results of this paper showed that, in comparison with other baseline methods (e.g., Support Vector Machine or K-Nearest Neighbors), SdA drives to about 14% more accuracy in noisy situations.

Another work that involved SdA was of Yan et al. [30]. They performed the detection of abnormal actions of combustion gas turbines by applying extreme learning
machines joined with SdA. Their results showed that the features detected by SdA led to a more improved classification in comparison with hand-crafted features.

Shao et al. [31] extracted features in a fault diagnosis system for rotating devices with the input of vibration data by applying Deep Neural Networks. The authors combined Denoising Auto-Encoders with Contractive Auto-Encoders. To diagnose the fault, they refined the learned features using Locality Preserving Projection, then put them into a softmax classifier. Seven conditions were considered in their system, including: rubbing fault, compound faults (rub and unbalance), four levels of imbalance faults as well as normal operation. The device status is identified based on exploitation of vibration data by the diagnosis system. It figures out whether the device is in fault or normal condition. Their approach used on the experiments to diagnose the fault of locomotive bearing devices and rotors showed that it could beat Convolutional Neural Network and other shallow learning methods.

Lee [32] supported that the detection of faults belongs to several defect types often appearing on headlight modules of cars in a setting of vehicle manufacturer by proposing a Deep Belief Network (DBN) model together with a cloud platform and an IoT deployment. The results showed that the DBN model outperformed two other baseline methods (i.e., Radial Basis Function, and Support Vector Machine) with regard to error rate in test datasets.

1.4 CONCLUSION

In this chapter, we have reviewed two promising technologies for Industry 4.0, namely BDA and ML. We focused on the data aspect of smart manufacturing, which is fast and massive, and cannot be handled efficiently by conventional approaches. Indeed, by employing BDA and ML, especially DL, a wide range of industrial applications is proven to be accelerated. Although few successful works were reported in the literature, we believe that an optimizing and fully automated production on a large scale could be achieved in the very near future because of these potential advanced technologies.

ACKNOWLEDGMENTS

This work has been partly supported by Vietnam National University, Hanoi (VNU), under Project No. QG.20.55.

REFERENCES

5. J. Han, J. Pei, and M. Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.


Big Data Analytics and Machine Learning for Industry 4.0: An Overview


J. Han, J. Pei, and M. Kamber. Data mining: concepts and techniques. Elsevier, 2011.


Impact of Blockchain-Based Cyber Security Implementing Industry 4.0


Hybrid Huffman Tree Model for Securing Medical Data Using Steganography


doi:10.1109/TCSVT.2020.2974884.


**Energy Conservation in IOT-Based Intelligent Transport System**


Khanna, Abhirup. and Rishi. Anand . 8IT based smart parking system.İh 2016 International Conference on Internet of Things and Applications (IOTA), pp. 266□□□0. IEEE, 2016.


Digital Information System for Attack Reduction in Industry 4.0


Maniaci, Nick., Building of a Knowledge Center, PMI Leadership Meeting in Toronto, Canada, 2005.


79 Hendricks, Boyd., The Role of the Knowledge Centre in Knowledge Management, http://managaementsite.com


An Intelligent Blockchain-Based Framework for Securing Smart Grid


---


Sun, http://blogs.sun.com/gbrunett/entry/amazon_s3_silent_data_corruption


Intelligent Fog Computing for Industrial Wireless Sensor Networks


Maniaci, Nick., Building of a Knowledge Center, PMI Leadership Meeting in Toronto, Canada, 2005.


Hendriks, Boyd., The Role of the Knowledge Centre in Knowledge Management, http://manageamentsite.com


Sun, http://blogs.sun.com/gbrunett/entry/amazon_s3_silent_data_corruption


A Secure Data Sharing Scheme Based on Blockchain for Industrial Internet of Things Using Consensus Algorithm


Smart Trust Management Scheme for Detecting Attacks in IoT and Ubiquitous Computing


IoT-Based Smart Pipeline Leakage Detecting System for Petroleum Industries

http://indiatoday.intoday.in/site/Story/77111/Aspire/Thegreatrevival.html.


IOT-Based Smart Irrigation Systems


S. Idhate, A. R. Bilapatre and H. Kalbande, Dam Monitoring system using wireless sensor network, 04(04), 1767-1769.


L. Feng, and Y. Wang, Environmental effect of tidal gate and measures of disaster reduction - a case study of the Tidal gate on Yaojiag river, in 5th International Conference on Bioinformatics and Biomedical Engineering (ICBEE), 2011, pp. 181-185.


An Efficient Decentralized Medical Prescription Tracking Using Blockchain


Blockchain for Industry 4.0


An Effective E-Learning Mechanism to Meet the Learning Demand of Industry 4.0


