

# Survey on deep learning applied to predictive maintenance

Youssef Maher, Boujemaa Danouj

Department of Electrical and Mechanical Engineering, University Hassan 1, Morocco

---

## Article Info

### Article history:

Received Nov 25, 2019

Revised May 4, 2020

Accepted May 18, 2020

---

### Keywords:

Deep learning

Industry 4.0

Predictive maintenance

Prognosis health monitoring

Remaining useful life estimate

---

## ABSTRACT

Prognosis health monitoring (PHM) plays an increasingly important role in the management of machines and manufactured products in today's industry, and deep learning plays an important part by establishing the optimal predictive maintenance policy. However, traditional learning methods such as unsupervised and supervised learning with standard architectures face numerous problems when exploiting existing data. Therefore, in this essay, we review the significant improvements in deep learning made by researchers over the last 3 years in solving these difficulties. We note that researchers are striving to achieve optimal performance in estimating the remaining useful life (RUL) of machine health by optimizing each step from data to predictive diagnostics. Specifically, we outline the challenges at each level with the type of improvement that has been made, and we feel that this is an opportunity to try to select a state-of-the-art architecture that incorporates these changes so each researcher can compare with his or her model. In addition, post-RUL reasoning and the use of distributed computing with cloud technology is presented, which will potentially improve the classification accuracy in maintenance activities. Deep learning will undoubtedly prove to have a major impact in upgrading companies at the lowest cost in the new industrial revolution, Industry 4.0.

Copyright © 2020 Institute of Advanced Engineering and Science.  
All rights reserved.

---

## Corresponding Author:

Youssef Maher,  
Department of Electrical and Mechanical Engineering,  
University Hassan 1,  
577, Route de Casa -, Settati, Morocco.  
Email: maher-youssef@hotmail.com

---

## 1. INTRODUCTION

According to Accenture, a renowned global management consulting and professional services firm [1], predictive maintenance could save up to 12% on scheduled repairs, reduce maintenance costs by up to 30% and eliminate up to 70% of failures. For example, a study by the national science foundation (NSF) indicates that the center for intelligent maintenance systems (IMS), which is a leading research center in the field of prognosis health monitoring (PHM), has generated an economic impact of more than \$855 million through the deployment of PHM technologies to achieve near-zero unplanned downtime and more optimized maintenance [2].

According to the International Organization for Standardization, "prognosis is the estimation of the failure time and the risk for one or more existing and future failure modes" [3]. There are many predictive maintenance standards, such as the machine information management alliance (MIMOSA), which has 7 modules that are used by the US Army. Figure 1 below describes the different prognostic steps in order monitor continuously the state of the system with the help of the distributed computing technology.

Deep learning is represented by a compilation of machine learning algorithms that can model high levels of abstraction from large amounts of data through the use of multilayer architectures. This technique has made considerable progress in several areas such as image classification, speech recognition, instant and reliable language translation [4] and even in the search for new elements in particle physics.

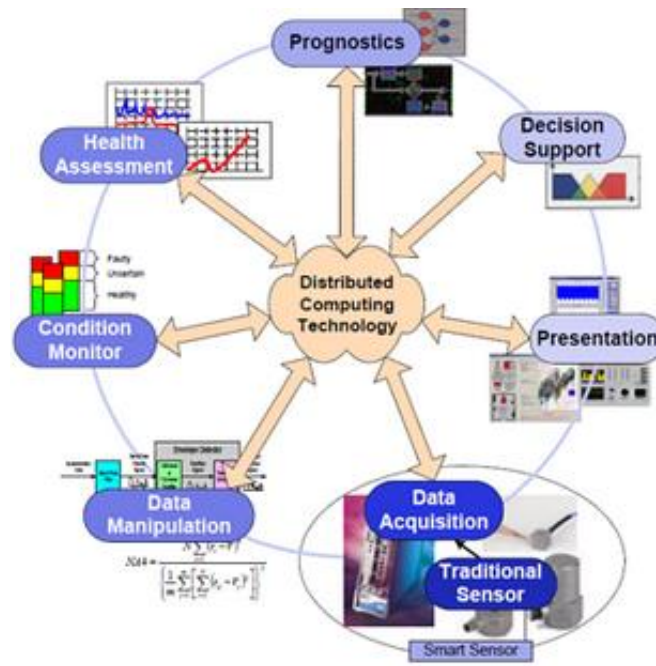


Figure 1. Prognostic steps [5]

This paper is structured as follows, Section 1 describes the limits of RUL estimation and the new advances in deep architectures to estimate a realistic RUL. Section 2 describes the new state of the art learning approaches and shows new opportunities for using cloud computing in deep learning. Section 3 summarizes the conclusions.

## 2. APPROXIMATION OF THE RUL BY VARIOUS ARCHITECTURES

The residual useful life (RUL) of a machine or component is the prediction of the time after which a component or a system will no longer be able to meet its operational needs after the observation of the first failures or alarms triggered. One of the goals of predictive maintenance is to obtain an estimated RUL that is as realistic as possible. Deep learning is the best tool to perform this task. This section provides an overview of this area of research.

### 2.1. Limits of RUL estimation by current methods

The decision making is based on the RUL confidence limits rather than on a single value. RUL prediction is difficult due to several important challenges, namely:

- Real systems are complex, and their behaviors are often nonlinear and nonstationary;
- A component may have different degradation curves due to different failure modes, resulting in different RULs (e.g., bearing cracks can occur in the inner ring, in the outer ring or in the cage, and each has its own degradation curve);
- The times required to achieve the same level of degradation by machines with the same specifications are often different;
- There is sometimes a complex temporal dependency between sensors; for example, a change in one sensor may cause a change in another sensor after a delay ranging from a few seconds to hours;
- The critical asset degradation history is sometimes nonexistent, such as a cooling engine in a nuclear power plant; in these cases, maintenance consists of regular replacement regardless of the actual conditions of the assets;
- When restarting newly installed assets, it will take a long time to collect pass-through data to accurately model the degradation; and
- The characteristics must have a good monotonic tendency to be well correlated with the fault propagation process to accurately approximate the RUL. In contrast, extracted entities that tend to have only dramatic changes near the end of the asset life are not exploitable. Figure 2 below shows the difference between good and bad RUL prediction.

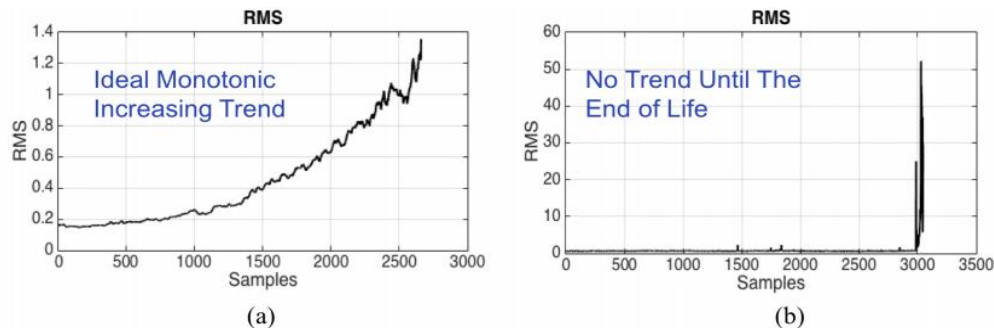


Figure 2. Characteristics of RUL prediction: (a) ideal, (b) not ideal [6]

## 2.2. Solutions provided by the deep learning

To address these uncertainties, prediction methods for the RUL based on deep learning have been successfully tested and are described as follows:

### 2.2.1. Deep belief network (DBN) and restricted boltzmann machine (RBM)

In [6] use an improved RBM with a new regularization technique that generates characteristics that are then employed as RBM input data. Finally, the RBMs are coupled with a self-organizing map (SOM) to improve the precision of the RUL prediction. Current methods prevent failure detection only when the end of life of the equipment is very close. In [7] propose a new approach, the multiobjective DBN ensemble, which allows for a compromise between accuracy and diversity by establishing an overall model for RUL estimation with outstanding performance.

### 2.2.2. Convolutional neural network (CNN)

In [8] used a novel deep architecture CNN-based regressor to estimate the RUL by employing the convolution and pooling layers to capture the salient patterns of the sensor signals at different time scales, unifying them and finally mapping them into the model. The resulting RUL estimation is efficient and accurate.

### 2.2.3. Variational auto-encoder (VAE)

In [9] described a semisupervised learning approach to predict asset failures when relatively little label information is available. The approach uses the nonlinear embedding-based VAE as a deep generative model, which is easy to train. The VAE was trained following the unsupervised learning process while utilizing all available data, both labeled and unlabeled. With this approach, the prediction accuracy was very high even with extremely limited available label information.

### 2.2.4. Recurrent neural network (RNN)

In [10] present an RNN HI for RUL prediction in a set of experimental data for accelerated bearing degradation and SCADA data of wind turbine degradation. The construction process was composed of three steps:

- Extraction of 14 characteristics: 6 characteristics of related similarity (1 temporal and 5 frequency) and 8 time-frequency functions are combined together;
- Selection of sensitive features; monotonic and correlation metrics select the most sensitive failure characteristics; and
- The building of the RNN-HI: the selected features are merged into one HI (RNN-HI) via an RNN. They used an LSTM network to solve the problem of rapidly increasing or vanishing gradients and a double exponential model to validate the effectiveness of the proposed RNN-HI approach.

The results show that the RNN-HI obtains relatively high monotonic and correlation values and better performance in RUL prediction than with a SOM HI method. In [11] use the RNN encoder-decoder (RNN-ED) with unsupervised learning on a set of aircraft engine and pump data. The RNN encoder extracts the important patterns in the time series subsequences of the entire operational life of the machines. The RNN decoder rebuilds the normal behavior, but it does not work well for the reconstruction of abnormal behavior. The trajectories of 20 engines are truncated at five locations to obtain five different instances. The RNN then maps the sensor readings on a health index (HI) trend curve to calculate the weighted average of the RUL approximations from the failed instances and then obtains the latest RUL estimation. The RNN-ED handles the reading of noisy sensors, missing data and the lack of previous knowledge about the degradation trends.

### 2.2.5. Long short-term memory (LSTM)

In [12] proposed using the LSTM for NASA dual-flow condition monitoring and achieved high performance in fault diagnosis and prediction under complex working conditions, mixed defects and noisy environments. The standard LSTM showed improved performance over the RNN by providing accurate information about the RUL for each fault simultaneously and also about the probability of occurrence of defects under complex operational modes and multiple degradations.

In [13] propose an unsupervised technique for reconstructing multivariate time series corresponding to normal behavior to obtain a health index (HI). To estimate the RUL, they implement an LSTM encoder-decoder (LSTM-ED) that works as follows: an LSTM encoder is used to map a multivariate input sequence to build a fixed-dimensional vector representation, and the LSTM decoder then uses this vector representation to produce the target sequence. The LSTM-ED-based HI model predicts future time-series, uses the prediction errors to estimate the health or novelty of a point and finally utilizes curve matching to estimate the RUL.

## 3. NEW STATE OF THE ART DEEP LEARNING APPROACHES

The main advantage of deep learning is its flexibility, which provides opportunities for improvement. Enhancements by using new learning approaches, new types of architectures, and new computing networks ensure continuous improvement and opportunities for deep learning at all levels.

### 3.1. Transfer learning

The transfer learning abilities that are learned and accumulated during previous tasks improve the performance of other neural networks when they are applied to a new task with a reduced training dataset. Traditional transfer learning methods assign the first  $n$  layers of a well-formed network to the target network, while the last layers in the target network are left untrained. They are then trained using the learning data of the new task.

In [14] use infrared thermal image transfer learning in the detection of machine defects and also in the prediction of the oil level. They use a modified VGG network (neural network created by the Visual Geometry Group at Oxford University), which is the best network for image data. The standard VGG is a very deep CNN with 16 layers with linear activation functions in each layer except the last one, which is a fully connected layer with a softmax activation function. The last layer is replaced by a new fully connected layer with a lower weight and fewer classes to accommodate learning transfer. Valuable information on important regions of thermal images can be obtained by applying Zeiler's method to this new VGG. This leads to a new level of understanding of the physical field.

In [15] propose three deep transfer strategies (DTL methods) based on an SAE autoencoder: weight transfer, transfer learning of characteristics and weight update. An SAE network is first trained with the failure data history of a cutting tool in an offline process, and this network is then employed with a new tool for an online RUL prediction. The DTL offers the possibility of extracting the characteristics from historical fault data, adapting and transferring them to a new tool and ultimately providing an effective RUL prediction with limited historical fault data.

In [16] present a transfer learning approach based on a pretrained deep convolutional neural network that is used to automatically extract input characteristics. They then go through a fully connected step to classify the characteristics obtained using experimental data composed of gear defects. The deep CNN network adopted as the basic architecture for Alexnet (five convolutional layers and three fully connected layers) in this study was originally proposed by [17] for object recognition (source domain) with ImageNet ILSVRC-2012 (1000 object classes and  $> 1$  million images). The classification accuracy of the proposed approach [18] exceeds those of other methods, such as the locally formed convolutional neural network and SVM. The precision obtained indicates that this approach is not only robust but can also be applied to fault diagnosis in other systems.

In [19] propose an unsupervised domain adaptation that does not require any label information from the target domain by only modifying the statistics of the BN layer. This AdaBN model gives high-level generalization ability to the DNN by transferring learned features from a source domain to the target domain without fine-tuning or additional components. Goodfellow I, Bengio Y and Courville A [20] describe a compromise between 'bias and variance' with beneficial generalization properties. Maximizing the prediction involves optimizing the model by incorporating external information.

### 3.2. Multimodal learning

In connection with transfer learning, multimodal learning can identify features that describe common concepts from different input types:

- With large unlabeled data, deep graphical models such as DBNs can be pretrained in an unsupervised way and then adjusted to a smaller number of labeled data because they learn a joint probability distribution from the inputs; and
- CNNs are mostly used with labeled data because they can be trained from end to end with a backpropagation function and demonstrate peak performance in many diverse tasks.

In [21] propose a novel multimode fault classification method based on DNN that resolves the problem of the load, mode and changing environment in which the machinery equipment operates. It is a hierarchical DNN model composed of three parts: the first hierarchy used for the mode partition, the second comprising a set of DNNs, which is devised to extract features of different modes separately and diagnose the fault source, and the third, which consists of another set of DNNs designed to distinguish the severity of the fault in a given mode. This approach allows for mode partitioning, which helps in the predictive maintenance of machinery and equipment.

### 3.3. Multitask learning

Multitask learning is complementary to multimodal and transfer learning. In the movie Karate Kid (1984), Mr. Miyagi teaches the child to do things that are not related to karate, such as sanding the floor and waxing a car. In hindsight, these kinds of tasks prove to be invaluable for acquiring karate skills. The purpose of an auxiliary task in MTL is to allow the model to learn representations that are shared or useful for the main task [22]. They use what they call cross-stitch units to allow the model to determine how task-specific networks exploit knowledge of another task by learning a linear combination of the previous layer outputs.

In [23] conclude that in multitask learning, success depends largely on "the similarity of the source semantics and target datasets." When only small amounts of target data are available, the sharing of concrete parameters can be considered as learning with a mean constraint, in which parts of all models (usually hidden layers) are forced to be the same as the average. The multitask learning architecture used is a bidirectional LSTM that consists of a single hidden layer of 100 dimensions shared between 10 word-processing tasks.

### 3.4. Deep learning optimization by cloud computing

Cloud computing could play a key role in scaling up deep learning. However, the main weakness is that recording the data from an increasing number of pieces of equipment and sending the data directly to the cloud creates a problem of system overload, inducing problems with speed, cost and security. One way to address this weakness is fog computing technology.

Fog computing is a decentralized computing infrastructure in which data, computation, storage and applications are distributed in the most efficient and logical place between the data source and the cloud. This technology reduces the resources allocated and keeps the bandwidth network at normal operating conditions, especially for deep learning. In fact, the inference is not done at the cloud level but at the fog level with sufficient memory and calculation resources, and the fog level will only transfer the necessary data to the cloud, such as the weight updates and the extracted features. The researchers in [24] proposed a new and efficient way to exploit deep learning by using fog-cloud computing and the capabilities of intelligent devices (edge devices). Figure 3 below describes the fog computing architecture connecting the different network levels and displays the computation, storage and local communication between edge devices. Figure 4 below shows the advantages of using the fog computing in terms of security, energy and speed.

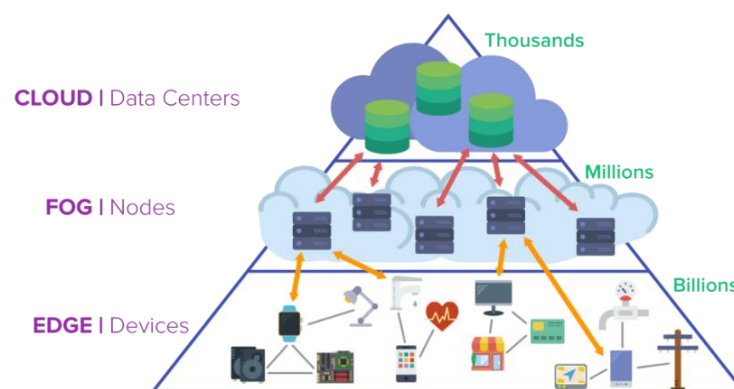


Figure 3. The cloud-fog technology and the interactions between levels [25]



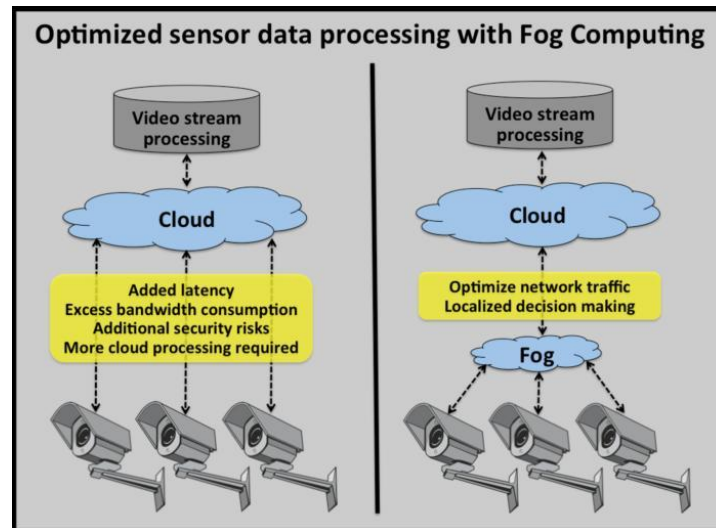


Figure 4. Optimizing the sensor network with fog computing [26]

The authors used a hybrid approach that combines the compression properties of neural networks such as CNN, RNN and LSTM and the incorporation of fog computing made possible by the integration of the increasing numbers of CPUs and GPUs in devices. The use of the fog in support of deep learning improves the user experience and optimizes the exploitation of the cloud. The fog embedded in the system will absorb the real-time data from the sensors and act quickly through the actuators. This technique is applicable to the field of health machine prognosis using the CNN for image classification of the health condition of machines by increasing the precision by 5% compared to traditional methods and saving data, energy and traffic, which makes it robust for applications in the new internet of things (IoT) industry.

#### 4. CONCLUSION

To quickly integrate the Industry 4.0 revolution, the PHM system needs to adopt established and high-performance predictive maintenance standards, such as the US Army's MIMOSA norm, to ensure optimal machine monitoring and high-quality manufactured products. This will ensure complete machine monitoring through well-structured communication systems and protocols, giving an edge in the highly competitive economic environment. With the development of deep learning, maintenance will become more efficient and more reactive in the near future and fit into the vision of Industry 4.0 because deep learning, with its multiple advanced architectures, eliminates the complexity of data analysis and the need for expertise and excessive work. That said, developing a so-called "intelligent" system in real time is essential to make companies interested by choosing to frame the studied system according to costs and benefits because the full potential of DL has not yet been explored.

#### REFERENCES

- [1] P. Daugherty, et al., "Driving Unconventional Growth through the Industrial Internet of Things," *Accenture Technology*, 2015.
- [2] Gray, D., Rivers, and G. Vermont, "Measuring the Economic Impacts of the NSF Industry/University Cooperative Research Centers Program: A Feasibility Study," Final Report, *Industry-University Cooperative Research Center Program National Science Foundation Arlington*, 2019.
- [3] Tobon-Mejia, D.A., K. Medjaher, and N. Zerhouni, "The ISO 13381-1 standard's failure prognostics process through an example," in *2010 Prognostics and System Health Management Conference*, 2010.
- [4] Wu, Y., et al., "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation," *arXiv preprint arXiv:1609.08144*, 2016.
- [5] Lebold, M. and K. Reichard, "OSA-CBM Architecture Development with Emphasis on XML Implementations," *Maintenance And Reliability Conference*, 2002.
- [6] Liao, L., W. Jin, and R. Pavel, "Enhanced Restricted Boltzmann Machine With Prognosability Regularization for Prognostics and Health Assessment," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7076-7083, 2016.
- [7] Zhang, C., et al., "Multiobjective Deep Belief Networks Ensemble for Remaining Useful Life Estimation in Prognostics," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2306-2318, 2017.

- [8] Sateesh Babu, G., P. Zhao, and X.-L. Li, "Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life," *Database Systems for Advanced Applications, Cham: Springer International Publishing*, pp. 214-228, 2016.
- [9] S. Yoon, A., et al., "Semi-supervised Learning with Deep Generative Models for Asset Failure Prediction," *arXiv preprint arXiv:1709.00845*, 2017.
- [10] Guo, L., et al., "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98-109, 2017.
- [11] Gugulothu, N., "Predicting Remaining Useful Life using Time Series Embeddings based on Recurrent Neural Networks," *arXiv preprint arXiv:1709.01073*, 2017.
- [12] Yuan, M., Y. Wu, and L. Lin, "Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network," *IEEE International Conference on Aircraft Utility Systems (AUS), China*, pp. 135-140, 2016.
- [13] Malhotra, P., et al., "Multi-Sensor Prognostics using an Unsupervised Health Index based on LSTM Encoder-Decoder," *arXiv preprint arXiv:1608.06154*, 2016.
- [14] Janssens, O., et al., "Deep Learning for Infrared Thermal Image Based Machine Health Monitoring," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 151-159, 2018.
- [15] Sun, C., et al., "Deep Transfer Learning Based on Sparse Auto-encoder for Remaining Useful Life Prediction of Tool in Manufacturing," *IEEE Transactions on Industrial Informatics*, pp. 1-1, 2018.
- [16] Cao, P., S. Zhang, and J. Tang, "Preprocessing-Free Gear Fault Diagnosis Using Small Datasets With Deep Convolutional Neural Network-Based Transfer Learning," *IEEE Access*, vol. 6, pp. 26241-26253, 2018.
- [17] Krizhevsky, A., I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in neural information processing systems*, vol. 25, pp. 1097-1105, 2012.
- [18] Liu, J., Y. Wang, and Y. Qiao, "Sparse Deep Transfer Learning for Convolutional Neural Network," *Thirty-First AAAI Conference on Artificial Intelligence*, pp. 2245-2251, 2017.
- [19] Li, Y., et al., "Adaptive Batch Normalization for practical domain adaptation," *Pattern Recognition*, vol. 80, pp. 109-117, 2018.
- [20] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville., "Deep Learning," *MIT Press*, 2016.
- [21] Zhou, F., Y. Gao, and C. Wen, "A Novel Multimode Fault Classification Method Based on Deep Learning," *Journal of Control Science and Engineering*, vol. 2017, no. 1, pp. 1-14, 2017.
- [22] Ruder, S., "An Overview of Multi-Task Learning in Deep Neural Networks," *arXiv preprint arXiv:1706.05098*, 2017.
- [23] Bingel, J. and A. Sogaard, "Identifying beneficial task relations for multi-task learning in deep neural networks," *arXiv preprint arXiv:1702.08303*, 2017.
- [24] Bin Qaisar, S. and M. Usman, "Fog Networking for Machine Health Prognosis: A Deep Learning Perspective," *International Conference on Computational Science and Its Applications*, pp. 212-219, 2017.
- [25] ERPINNEWS, "Fog computing vs edge computing," 2018. [Online], Available: <https://erpinnews.com/fog-computing-vs-edge-computing>.
- [26] Chelliah, P.R. and J. Pushpa, "Expounding the edge/fog computing infrastructures for data science, in cloud and fog computing infrastructures for data science," *IGI Global*, pp. 1-32, 2018.

## BIOGRAPHIES OF AUTHORS



**Youssef Maher** is a research scholar in the Department of Electrical and Mechanical Engineering, from the University Hassan 1, Settat, Morocco. He has a degree in industrial and mechanical engineering. His research interests include Data Mining, Big data analytics, Deep Learning, Neural Networks Predictive analytics.



**Boujemaa Danouj** Doctor Engineer in mechanical engineering. Graduate of the National School of Electricity and Mechanics Casablanca, Morocco. Master of Philosophy (M.Phil) from Middlesex University, UK. Ph.D. from ÉTS (École de Technologie Supérieure), Montréal, Canada. Expertises: Free form curves and surfaces modeling, partial discharges in high voltage machines, deep learning and predictive maintenance, Metrology and tolerancing.