

Design and Implementation of Missing Data Classification Technique for IoT Applications Using Artificial Intelligence

Gopal Patil^a, Dr. Raj Thaneeghaivel V^b

^a (Ph.D. Research Scholar), ^b (Associate Professor)

^{a,b} Department of Computer Science and Application, Sarvepalli Radhakrishnan University, NH 12, RKDF IST Campus, Hoshangabad Road, Misrod, Bhopal (M.P.)

Article History: Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

Abstract: The combination of various sensors with different data methods is a common technique used to increase precision in the classification of IoT health data. However, for even the assessment outcomes, all modalities are barely available and this scarcity of evidence poses significant barriers to multimodal education. Driven by recent developments in deep education, we are providing a cross-neural network for the segmentation of the IoT Health Data Classification, which is trained on data modalities not all available during trials. In IoT Health Data Classification, we train our architecture with a cost function that is especially tailored to unbalanced classes. We are providing the device with a benchmark data set with incomplete data. Assuming that they are not present in the research process, our methodology goes beyond both the CNN training and the collection of two CNNs trained in the missing modality by utilising time data

Keywords: Training , Data integration , Data models , Cross Neural Network

1. Introduction

As the automated network system of the Internet of Things (IoT) grows and evolves, IoT models become complicated on a regular basis[1],[2]. People are happy with a data-driven architecture that leads research to machine learning apps alongside IoT. IoT and Deep Learning approaches are currently utilised in all areas of human life. The simulation of brain impulses entails the application of artificial learning approaches[3] in medicine, ECG interpretation, X-ray disease recognition, genetic sequence detection, and an automated pathology tool for carcinogen detectives. Machine learning methods may also be seen in the aerospace industry. D'Angelo et al.[4] applied content-based image recovery and machine learning strategies to an electrical impedance aeroplane generated by current eddy experiments. The latest Eddy test is a complex task used to detect defects in the aircraft industry.

Apart from deep learning, IoT tools are also applied to these domains. The increasing complexity of IoT infrastructure increases its undesirable vulnerability. Data breaches and anomalies in IoT applications have been common. IoT equipment uses wireless media to transmit data, making it easier to attack[5]. Typical local networking assault is confined to local or small local nodes, but the IoT attack spreads through a larger area and has a devastating effect on IoT sites[6]. A secure IoT infrastructure is needed to defend against cybercrime. With the vulnerability of the IoT gui, the security measures used became vulnerable. Data is the company's capital for many owners and founders. Any of the records is limited and exclusive to the government and some private companies. IoT node vulnerability causes confidential information to be obtained by every significant entity[7] at the back of the attacker. As mentioned above, there are several trivial approaches to solve the problems. Attacks and irregularities are stored in a database utilising a signature-based[8] method.

In addition, this unit is tested against the database at specific intervals. However, this development technique generates an overhead and is vulnerable to unforeseen threats. The advantage of data processing technology is that it works faster and can overcome the problem posed by uncertain threats. Accordingly, this article integrates data analysis processes. The primary objective of the system is to create an efficient, reliable and successful IoT architecture that can identify its failure, protect the firewall against cyber attacks and recover automatically. The machine-based learning solution that can recognise and protect a system in an irregular situation is proposed here. Several classifiers in deep learning have been used for this task. Another key point in this paper is that a simple model such as Decision Tree and Random Forest may be connected to a complex network such as ANN for the detection of anomalies.

2. Related work

Firouzi F. et al[1]The purpose of this research is to provide readers with a description of machine learning. Second, we should talk about the basic hypotheses of probability, statistics and linear algebra, as these are the fundamentals from which a lot of machine learning solutions can be supported. Next, IoT solutions offer us more instances of machine learning. Finally, the two key forms of machine learning, supervised learning and unattended teaching would be discussed.

Thakkar, A., et al[2] This paper on the Intrusion Detection System (IDS). offers a comprehensive IoT survey for the years 2015–2019In IoT architecture, we looked at various IDS placement and IDS technical strategies the analysis tackles several IoT intrusions. The paper also discusses risks to security and barriers to IoT.

Liu, Q., et al[3]The real-time BDA framework integrating the SLN network and the Deep Neural Network will reduce the burden from frequent MEC calculations on the Deep Network substantially, and reduce the MEC energy consumption for remote real-time monitoring. Real-time Large Data analysis system

Raeesi Vanani I., et al[4]Machine learning can evaluate and streamline the diagnosis method in a large spectrum of IoT device information. The literature focuses on machine learning techniques for data on health devices for disease identification and prediction. The first purpose of this chapter is to clarify the approaches to machine learning and integrated solutions to IoT data for disease detection. This chapter deals with the past of machine learning and a variety of important and functional machine learning algorithms in the field of healthcare.

Keserwani, P.K., et al[5]The Internet of Things (IoT) applies technological creativity to the development of informative environments to facilitate the position of people. Technological advances provide businesses with many means of detecting and exploiting various attacks that may prevent the security of IoT networks. The key issue of the IoT network model is therefore security and privacy. Machine and IoT networks need to be secured against various forms of threats and dangers.

Efat M.I.A., et al[6]Based on the patient's form and previous health history, the range of risk status can differ. In addition, an automatic phone call and/or SMS response is made to the relative and location of the patient, if he/she has a mild to severe health threat. In comparison, where there is a significant risk, patients are named closest to the hospital.

Jabeen, F., et al.[7]In the second segment, the cardiac patient is recommended by age and gender for the physical and nutritional regime. Professional cardiologists can be used to gather data the performance of the system is measured and 98 percent is achieved.

3. Proposed Methodology

Each participant could contribute to any given combination of S sources (here $S=4$: tapping, walking, voice, and memory) so the database includes a possible combination of 2^S of available tests. The $I[1...S]$ binary vector can be represented for each different combination of sources where $I[i]=1$ demonstrates a contribution to i th. A participant is assigned a binary source vector based on the sources contributed in this research. The vector of the binary source defines a participant's domain. The domain assignment process for a mPower dataset participant is demonstrated [8]Then we divide the initial source dataset in smaller but complete fields, which is called a dataset deconstruction, by assigning the participant to a single domain. There is a degree of overlap in which sources between certain domains are available.

In Fig, for example. 1 It is clear that domains 7, 13 and 15 (with the exception of at least one other source) contain the same sources as those found in domain 5. A multi-task learning system (M-TL) can be used to share data across different areas and multiple learning tasks can be addressed at the same time[9]

The M-TL results are difficult to understand because of these confusing factors. [10] We note two special cases of data set deconstruction, in order to overcome these limitations. In the mPower datasets are 1, 2, 4 and 8 domains that are consistent with the storey, voice, walking and tapping of individual sources domains. The data set is compiled with the same source dofmain. the particular source model for each source domain, regardless of the domain to which they were assigned [11]

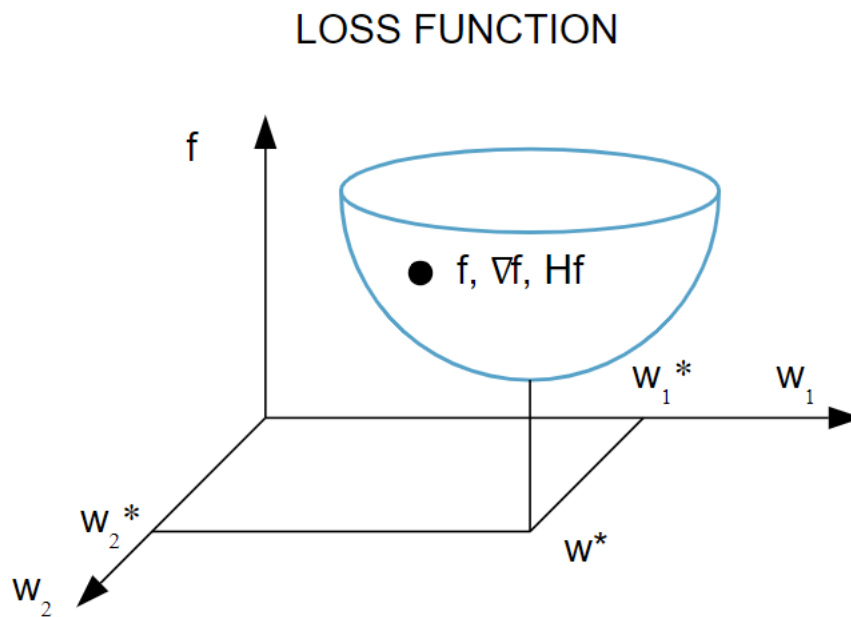
In addition, the individual source models can be merged into all possible combinations by using source ensembles. Participants with complete source data can also test all models created by means of source data sets. Now, our model framework can be formally defined. With the deconstruction of the source-wise missing dataset, all participants, including those with missing data, can develop individual S-source models.[12]

Comprehensive source data participants are excluded and reserved as a test kit for the testing of all individual sources and their 2S combinations from the training/validation of different source models. The results of each model are directly comparable to each other, eliminating the confusion factor mentioned above, after creating a consistent test set for all models. In the mPower dataset, we identified participants forming a training and validation set with incomplete source data. [13]The conjugate gradient method can be considered as an intermediate between the gradient descent and the Newton method. There is a guiding impulse for incremental acceleration, generally correlated with descent gradient convergence. This approach also prevents information from being required for calculating, storing and reversing the Hessian matrix using the Newton process.[14]

4. Learning problem

The problem of learning is formulated in terms of the index of inability to reduce, ff. It measures the performance of the neural network for data collection.[15]

In addition, the loss index is made up of error and regulatory terms. The error term tests how the data set integrates into a neural network. The regulatory concept is used to prevent overcasting by checking the complexity of the neural network.



The minimum loss function is located at the absolute point w , as can be seen in the previous image. At both points, AA, the first and second loss function derivatives can be evaluated.

The first derivatives are grouped in the gradient vector, whose elements can be written as

$$\nabla f(w) = \partial f / \partial w_i, \nabla f(w) = \partial f / \partial w_i,$$

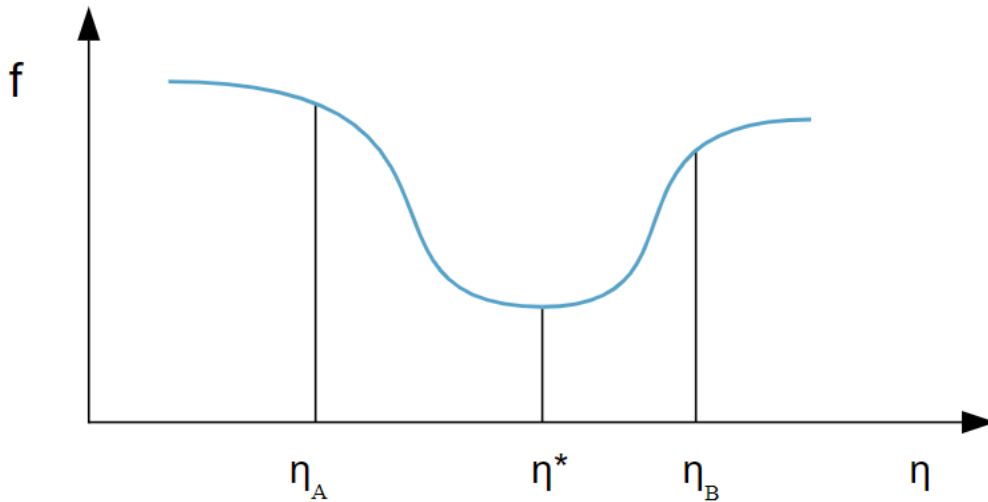
for $i=1, \dots, n; i=1, \dots, n$.

Likewise, the second Loss Function derivatives can be categorised into the Hessian matrix.

$$H_{i,j} f(w) = \partial^2 f / \partial w_i \partial w_j, H_{i,j} f(w) = \partial^2 f / \partial w_i \partial w_j,$$

for $i,j=0,1, \dots; i,j=0,1, \dots$

ONE DIMENSIONAL OPTIMIZATION



5. One-dimensional optimization

Since the loss function is subject to many variables, the usage of one-dimensional optimization techniques is especially important. In specific, they are widely used in the training process of the neural network.[16]

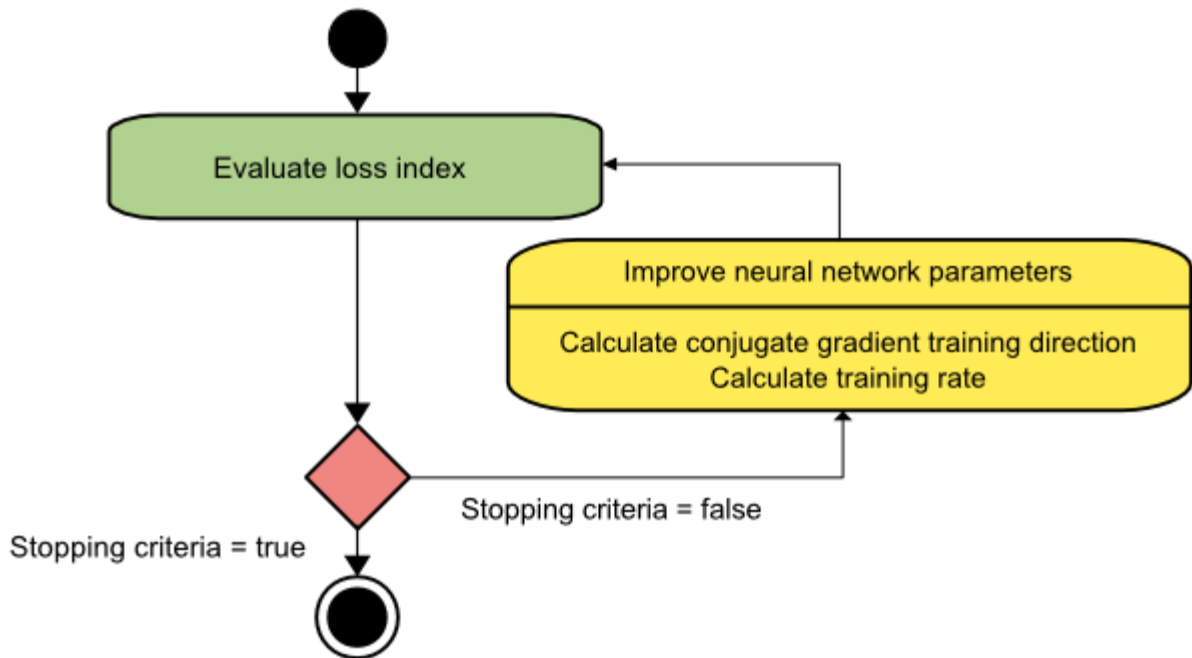
One-dimensional optimization techniques check the minimum one-dimensional function in this relationship. The Gold Segment Method and the Brent Method are some of the algorithms that are widely used. The minimum bracket is both limited to a width smaller than the tolerance between the two outer points of the bracket.[17]

The quest for conjugate directions is accomplished by a conjugate gradient training algorithm, which typically provides faster convergence than steep descent directions. These instructions are combined with the Hessian matrix.

Let's mark the direction of vector preparation. The conjugate gradient approach then generates a series of training directions such as $d(i+1)=g(i)$, $d(i+1)=g(i)$ in the following: $I(i)$. $I \mu g(i) = g(i)$ The conjugate gradient procedure builds the following exercise path sequence: (i) , for $i=0,1,\dots,i=0,1,\dots$

Here μg is referred to as the conjugate parameter and there are different ways of calculating it. Both Fletcher and Reeves, as well as Polak and Ribiere, are widely used. The training direction is periodically reset to the gradient negative for all conjugated gradient algorithms.[18]

for $i=0,1,\dots,i=0,1,\dots$ The accompanying example depicts an action diagram with a conjugate gradient for the preparation process. The parameters are improved here, first by calculating the conjugate direction of the gradient and then by providing the appropriate training rate to that end.



This method has been more effective than gradient descent in the training of neural networks. Since there is no need for a Hessian matrix, a conjugate gradient is also suggested for large neural networks.

CNNs are widely known as modern machine teaching approaches and, as a result of the generally small size of PD datasets, little attention has been extended to the PD classification to date[19]. CNNs do not need to define an explicit set of features, but may learn features or filters directly from raw data. In addition, these filters are translationally invariant and make CNNs especially suitable for noisy raw data, such as the mPower database. The tapping, walking and speaking movements of the four activities of the mPower database provide details ideal for use as CNN inputs. along with the pixel touch screen info. Due to the irregular sampling of the touch screen data, we use linear interpolations to generate waveforms of similar width to the waveforms of the accelerometer[20].

The raw input ($f_s=100$ Hz) has been used for triaxial accelerometers and triaxial gyroscopes. The raw speech signal ($f_s = 42$ kHz) was used for the operation of the sound. Per waveform was standardised for zero mean unit variance, except the interpolated touch-screen tapping form. The multi-channel architecture of CNN for both styles of operations. We favoured a standard general architecture, although additional advantages could be obtained by the use of different network architectures for each source. Here, two convolutionary divisions use the concept of the vector of the first receptive field width. The data frequency components are better captured utilising large-range convolutionary filters. The time of the signal is better recorded by using restricted width filters. Thus, the distance of the first convolutionary filter is different in each channel, such that both the transient and frequency components of the data are obtained. In order to capture data frequency elements, alternative CNN Architectures use thin, receptive areas, but need several more layers and convolutionary operational activities[21].

6. Results and Analysis

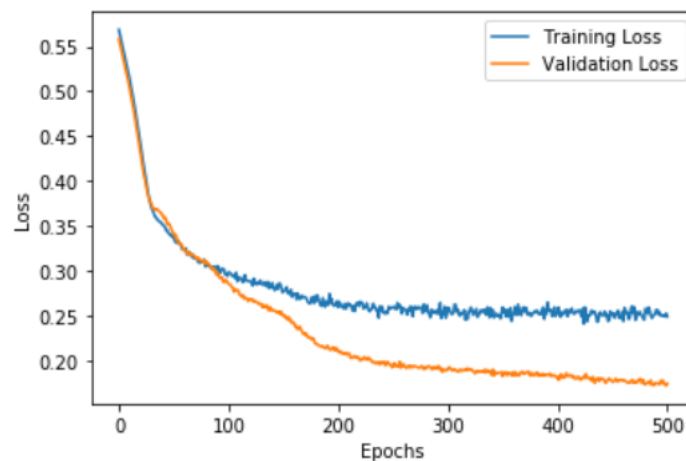


Figure 1: Loss data classification in term of training loss and validation loss

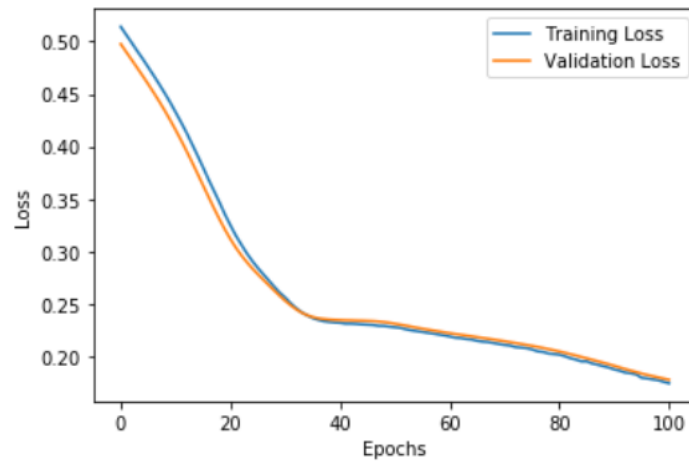


Figure 2 : Loss data classification in term of training loss and validation loss

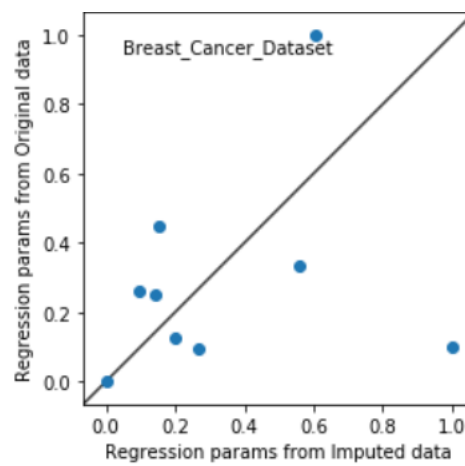


Figure 3 : Breast cancer data classification

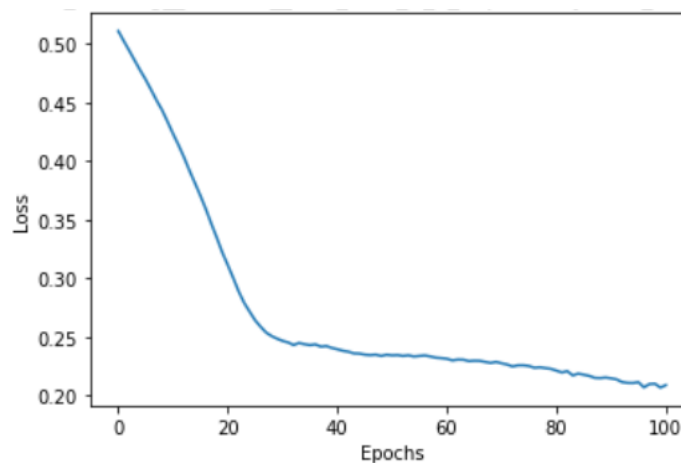


Figure 4 : Loss data classification

7. Conclusion

Two types of comparative research were undertaken to determine the feasibility of multi-source ensemble instruction. Second, we implemented the most common approach for data sets in which the source data is absent. The planning of the total data collection has been introduced. Training and evaluation models with complete data from the participants showed that the precision of the classification was enhanced in line with single source models. We also note that only 133 (8.8%) of the 1 513 participants participating in this study had complete

source results. As a result, 91.2% of participants are discarded when we use full data collection learning. This is the conventional approach of literature, and is definitely a very inefficient use of data[20]. The second comparative study involved an estimation of the effect of the selection of roles. During Incomplete data set learning, a significant number of participants with incomplete data were used to separately carry out the compilation of features for each source, and these features were used by participants with full classification. A single neuron model (LR) was used in the comparative studies to assess the functioning of the two systems. If more complicated models such as random forests and DNNs, their inherent capacity to choose features will change their rating accuracy. This filtering procedure has been found to be more efficient in classification than the entire approach of the data collection.

References

- Firouzi F., Farahani B., Ye F., Barzegari M. (2020) Machine Learning for IoT. In: Firouzi F., Chakrabarty K., Nassif S. (eds) *Intelligent Internet of Things*. Springer, Cham. https://doi.org/10.1007/978-3-030-30367-9_5.
- Thakkar, A., Lohiya, R. A Review on Machine Learning and Deep Learning Perspectives of IDS for IoT: Recent Updates, Security Issues, and Challenges. *Arch Computat Methods Eng* (2020). <https://doi.org/10.1007/s11831-020-09496-0>
- Liu, Q., Sun, S., Yuan, X. et al. Ambient backscatter communication-based smart 5G IoT network. *J Wireless Com Network* 2021, 34 (2021). <https://doi.org/10.1186/s13638-021-01917-3>.
- Raeesi Vanani I., Amirhosseini M. (2021) IoT-Based Diseases Prediction and Diagnosis System for Healthcare. In: Chakraborty C., Banerjee A., Kolekar M., Garg L., Chakraborty B. (eds) *Internet of Things for Healthcare Technologies*. Studies in Big Data, vol 73. Springer, Singapore. https://doi.org/10.1007/978-981-15-4112-4_2
- Keserwani, P.K., Govil, M.C., Pilli, E.S. et al. A smart anomaly-based intrusion detection system for the Internet of Things (IoT) network using GWO–PSO–RF model. *J Reliable Intell Environ* 7, 3–21 (2021). <https://doi.org/10.1007/s40860-020-00126-x>.
- Efat M.I.A., Rahman S., Rahman T. (2020) IoT Based Smart Health Monitoring System for Diabetes Patients Using Neural Network. In: Bhuiyan T., Rahman M., Ali M. (eds) *Cyber Security and Computer Science*. ICONCS 2020. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 325. Springer, Cham. https://doi.org/10.1007/978-3-030-52856-0_47.
- Jabeen, F., Maqsood, M., Ghazanfar, M.A. et al. An IoT based efficient hybrid recommender system for cardiovascular disease. *Peer-to-Peer Netw. Appl.* 12, 1263–1276 (2019). <https://doi.org/10.1007/s12083-019-00733-3>.
- Bhardwaj, S., Pandove, G. & Dahiya, P.K. An efficient comparison of two indexing-based deep learning models for the formation of a web-application based IoT-cloud network. *J Ambient Intell Human Comput* (2020). <https://doi.org/10.1007/s12652-020-02500-2>
- Hiriyannaiah S., Khan Z., Singh A., Siddesh G.M., Srinivasa K.G. (2020) Data Reduction Techniques in Fog Data Analytics for IoT Applications. In: Tanwar S. (eds) *Fog Data Analytics for IoT Applications*. Studies in Big Data, vol 76. Springer, Singapore. https://doi.org/10.1007/978-981-15-6044-6_12
- Hiramoto R.E., Haney M., Vakanski A., Shareef B. (2019) Toward a Secure IoT Architecture. In: Kondratenko Y., Chikrii A., Gubarev V., Kacprzyk J. (eds) *Advanced Control Techniques in Complex Engineering Systems: Theory and Applications*. Studies in Systems, Decision and Control, vol 203. Springer, Cham. https://doi.org/10.1007/978-3-030-21927-7_14.
- Shi Z., Li X., Su Z. (2018) Power Missing Data Filling Based on Improved k-Means Algorithm and RBF Neural Network. In: Sun X., Pan Z., Bertino E. (eds) *Cloud Computing and Security*. ICCCS 2018. Lecture Notes in Computer Science, vol 11067. Springer, Cham. https://doi.org/10.1007/978-3-030-00018-9_48
- A. Singh Rajawat and S. Jain, "Fusion Deep Learning Based on Back Propagation Neural Network for Personalization," 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 2020, pp. 1-7, doi: 10.1109/IDEA49133.2020.9170693.
- Deepa N., Prabadevi B. (2020) Advanced Machine Learning for Enterprise IoT Modeling. In: Haldorai A., Ramu A., Khan S. (eds) *Business Intelligence for Enterprise Internet of Things*. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-44407-5_5
- Mani J.J.S., Rani Kasireddy S. (2019) Population Classification upon Dietary Data Using Machine Learning Techniques with IoT and Big Data. In: *Social Network Forensics, Cyber Security, and Machine Learning*. SpringerBriefs in Applied Sciences and Technology. Springer, Singapore. https://doi.org/10.1007/978-981-13-1456-8_2
- A. S. Rajawat, O. Mohammed and P. Bedi, "FDLM: Fusion Deep Learning Model for Classifying Obstructive Sleep Apnea and Type 2 Diabetes," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2020, pp. 835-839, doi: 10.1109/I-SMAC49090.2020.9243553.

- A. S. Rajawat and A. R. Upadhyay, "Web Personalization Model Using Modified S3VM Algorithm For developing Recommendation Process," 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170701.
- Y. Shen, H. Zhang, Y. Fan, A. P. W. Lee and L. Xu, "Smart Health of Ultrasound Telemedicine Based on Deeply-Represented Semantic segmentation," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2020.3029957.
- J. Zhang, P. Liu, F. Zhang, H. Iwabuchi, A. A. d. H. e. A. de Moura and V. H. C. de Albuquerque, "Ensemble Meteorological Cloud Classification Meets Internet of Dependable and Controllable Things," in IEEE Internet of Things Journal, vol. 8, no. 5, pp. 3323-3330, 1 March1, 2021, doi: 10.1109/JIOT.2020.3043289.
- C. A. C. Montañez and W. Hurst, "A Machine Learning Approach for Detecting Unemployment Using the Smart Metering Infrastructure," in IEEE Access, vol. 8, pp. 22525-22536, 2020, doi: 10.1109/ACCESS.2020.2969468.
- A. Mongardi, P. M. Ros, F. Rossi, M. R. Roch, M. Martina and D. Demarchi, "A Low-Power Embedded System for Real-Time sEMG based Event-Driven Gesture Recognition," 2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS), Genoa, Italy, 2019, pp. 65-68, doi: 10.1109/ICECS46596.2019.8964944.
- A. Al-Abassi, H. Karimipour, A. Dehghantanha and R. M. Parizi, "An Ensemble Deep Learning-Based Cyber-Attack Detection in Industrial Control System," in IEEE Access, vol. 8, pp. 83965-83973, 2020, doi: 10.1109/ACCESS.2020.2992249.