International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

A survey on Machine Learning Algorithms for Wireless Internet of Things (IoT)

Ms. Priti V Shinde[1], Ms. Rupali R. Shewale[2]

[1,2]*Assistant Professor, Department of Computer Engineering, MVPS's KBTCOE, Savitribai Phule Pune University Nashik, Maharashtra, India*

Abstract - *In the present world of rapidly growing massive technologies and their inter-connection with hardware devices and software applications plays a important role for the appearance of different sensor devices that are inter-connected through internet in order to establish interaction with the physical objects in the world. Various hardware devices connected over the internet is defined as Internet of Things (IOT). IOT generates massive amount of data with respect to various characteristics and qualities of data. Machine learning combination with IOT gives the persistent development to extend the intelligence of the IOT devices and applications. The exposure of different smart IOT applications with machine learning helps in observation, systematic analysis, processing and smart usages of the large volume of data in different fields. Many industries are using the machine learning and more specifically by using IOT's potential. This paper consists of machine learning basic introduction, machine learning algorithms, overview of Internet of things, statistic of WSN model and Machine Learning algorithms. And the last section of the paper consists of conclusion.*

Key Words: **Internet of Things, Machine learning, Wireless Sensor Network, IoT, ML**

1. INTRODUCTION

Internet of Things, the term given by K. ASH-TON in 1999 is defined as to explain a network of interconnected devices such as sensors, actuators, mobile phones which interact and work together to attain common goal.In the next few years, cars, kitchen appliances, televisions, smart phones, utility meters, intra-body sensors, thermostats, and almost anything we can imagine will be accessible from anywhere on the planet [3].The revolution brought by the IoT has been compared to the building of roads and rail-roads during the Industrial Revolution of the 18th to 19th centuries [4] and is expected to drastically transform the education, health- care, smart home, manufacturing, mining, commerce, transportation, and surveillance fields, just to mention a few [5].As IoT gained many aspects of our lives, the wireless resources will also be in high demand.

According to the latest Ericsson's mobility report, there are now 5.2 billion mobile broadband subscriptions worldwide, generating more than 130 exabytes per month of wireless traffic [6] .Over 50 billion devices are expected to be in the IoT by 2020, which will generate a global network of "things" of dimensions never seen before [7].Given that only a few radio spectrum bands are available to wireless carriers [8], technologies such as radio frequency (RF) spectrum sharing through beam forming [9–11] will become essential in the near future. These technologies typically require synchronization among wireless devices to enhance spectrum usage. They need to be implemented in a distributed manner to ensure scalability, reduce overhead and energy consumption.

To report this challenge Machine Learning technologies are used. Various Machine Learning algorithms are present which are used in this context. Further sections focuses on them.

2. LITERATURE SURVEY

2.1 An overview of Machine Learning

It is the concept that systems can learn from present data, identify patterns and make decisions without human. Some of the Machine Learning application which are known very well are fraud detection, predictive maintenance, healthcare and etc. Machine Learning combines data with statistical tools to calculate an output. This output is then used to makes actions. The main action provided by Machine Learning is recommendation along with the important objective learning and inference. Very First, the machine learns through the discovery of patterns where data plays an important role. Hence selection of good data should be choosing carefully.

Machine Learning can be classified into supervised learning, unsupervised learning, reinforcement learning.

1. Supervised Learning

In supervised learning both input and output data are provided. Input and output data are labelled for classification. Hence supervised learning algorithms learn from a training dataset. An algorithm uses training data and feedback from us to learn the association of given inputs to a given output. For instance, a person can use marketing expense and weather forecast as input data to predict the-sales of fruit juice cans. We can use supervised learning when the output data is known. The supervised learning algorithm will predict new data. Supervised learning is a learning in which the machines are trained using data which is well labelled that means some data is already tagged with

the correct answer. After that, the machine is provided with a new set of data so that supervised learning algorithm analyses the training data (set of training examples) and gives a correct outcome from labelled data. The objective of supervised learning is to learn how to predict the appropriate output data for a given input data [25]. For example, *Amazon's Recommendation System, voice assistant, Weather Apps, Gmail Spam Filtration* etc. It also helps in prediction of future results for unseen data [26].

There are two categories of supervised learning:

- Classification task
- Regression task

1.1 Classification task

Suppose we want to predict the gender of students in class. We will start gathering data on the height, weight, percentage etc. from our student database. We know the gender of each of our student; it can only be male or female. The objective of the classifier will be to assign a probability of being a male or a female (i.e., the label) based on the information (i.e., features we have collected). When the system learned how to identify male or female, we can use new data to make a prediction. For instance, you just got new information from an unknown student, and we want to know if it is a male or female. If the classifier predicts male = 70%, it means the algorithm is sure at 70% that this student is a male, and 30% it is a female. Any label can be having two or more classes. In the above example it has only two classes, but if a classifier needs to predict object, it has many classes such as glass, table, shoes, etc. each object represents a class.

1.2 Regression task-

It is a Supervised Learning task where output is having continuous Real value or real value. Such as dollars, weight, temperature sensors values. Regression technique predicts a single output value using training data, for e.g. we can use regression technique to predict the price of a house from training data. The input variables will be area of location, size of a house.

2. Unsupervised Learning-

Unsupervised learning is where we only have input data and no corresponding output variables. These are called unsupervised learning because unlike supervised learning above there is no correct answer. Algorithm discovers the data and presents the interesting structure in the data. For example, *NASA* uses this learning approach to create the different clusters of heavenly bodies each of which consists of similar nature objects. [26].

Unsupervised learning problems are classified into clustering and association problems.

- Clustering: A clustering is where we want to discover the natural groupings in the data, such as grouping customers by purchasing behavior.
- Association: An association rule learning problem is where we want to discover a rule that describes large parts of data, such as people that buy product p also want to buy q.

3. Reinforcement learning-

It deals with the problem of learning the appropriate action or sequence of actions to be taken for a given situation in order to maximize payoff[25]. This paper focus is on supervised and unsupervised learning since they are being widely used in IoT smart data analysis.

2.2 Internet of Things(IoT)

The Internet is the largest heterogeneous network and infrastructure in existence. It connects many devices such as touch screen devices (smart phones, tablets) and which evolved advancement of wireless and mobile technologies. On the other hand, the Internet of Things (IoT) is a fast growing heterogeneous network of connected sensors and actuators attached to a wide variety of devices. Mobile and wireless technologies contributes to the connectivity and progress of IoT. The future will have smart and low-power

Table -1: Network sizes for WSN technologies

Technology	Network Size
ZigBee	Approximately up to 65,000 nodes
Bluetooth	Eight nodes per network/piconet
Wi-Fi (802.11a/ac)	2007 associated with an AP
Wi-Fi 802.11ah	Approximately 8000 nodes
Ly7oRa	LoRa gateway must have a very high capacity or capability to receive messages from a very high volume of End nodes

networked devices connecting to each other and to the Internet using reliable low-power wireless transmissions.

1. Wireless Low power technologies for IoT- Various wireless technologies are present such as IEEE 802.15.4 technologies, Bluetooth Low Energy, and Wi-Fi, IEEE 802.11ah and LoRa technologies. IEEE 802.11ah LoRaWAN are the latest technologies in long-range and low-power WAN. These technologies are targeted for low-power and low-cost devices.LoRa plays a major role in the future of wireless and machine to machine (M2M) communications. The 802.11 protocol with its 802.11a/b/g/n/ac variants is widely used for the IoT. Now a days, almost every where a Wi-Fi network is present. Wi-Fi has become the important term when referring to connecting to the Internet via a wireless access point. Many IoT devices require the use of a low-cost and low-power wireless technology when connecting to the Internet [6].In the evolvement of the IoT,

low power consumption, low cost technology, security, simplicity of technology are considered. ZigBee and Bluetooth are trying to provide these solutions as well as competing to provide these to customers.

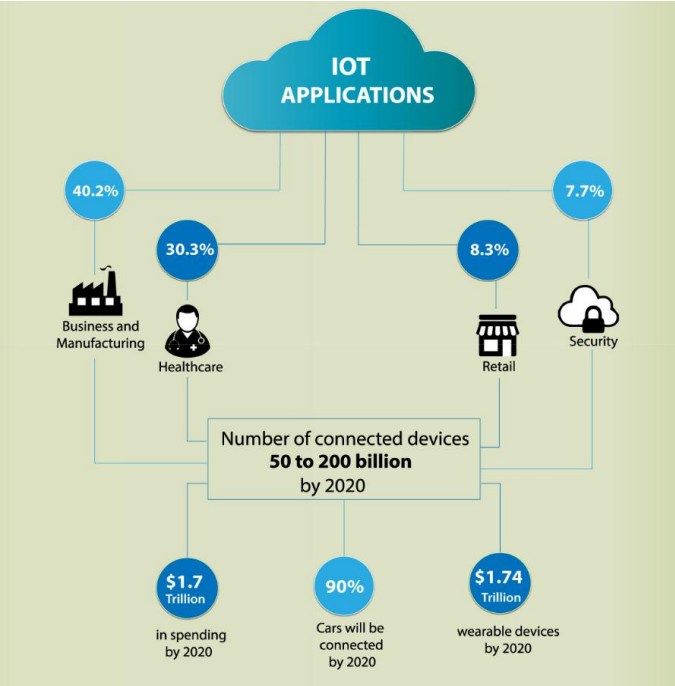

Figure 1- IoT Growth by 2020[1]

Cellular technology can contain a larger number of devices; the costs involved are significantly higher than those related with other technologies such as ZigBee. LoRa can support a large volume of devices.

Table -2: Comparison of WSN Technologies

Technology / Features	Wi-Fi	Blue-tooth	ZigBee	6LOWPAN	LoRa
Network	LAN	PAN	LAN	LAN	LAN
Topology	Star	Star	Mesh, Star, Tree	Mesh, Star	Star & star of star
Power	Low-High	Low	Very Low	Very Low	Very Low
Data Rate	Up to 1.3 GBPS	2.1 mbps	250 kbps	200 kbps	0.3 kbps-100 kbps
Range	Up to 100 m	< 100 m	10-20 m	10-20 m	3.5 km urban area
Network Size	Medium	Small	Very large	Very large	Medium large

The communication protocols at various layers of the IoT protocol stack are shown in Fig.2

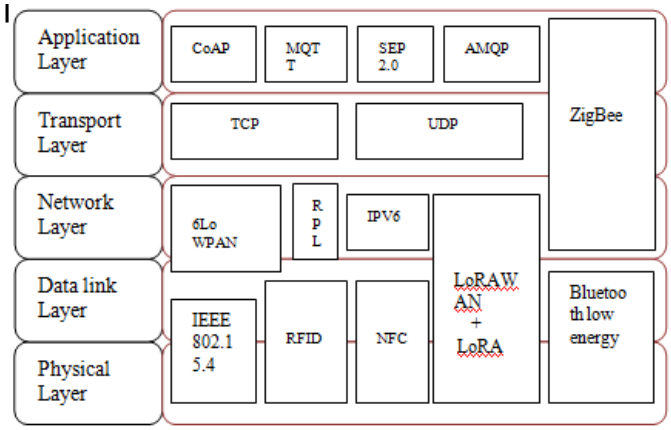

Figure 2- WSN IoT Protocol stack

3. STATISTICS OF WIRELESS IOT MODELS BASED ON ML ALGORITHMS

Every time the IoT sensors gather data, there should to be someone at the backend to classify the data, process them and it has to be ensured that information is sent out back to the device for decision making.

Each of the machine learning class has many types of algorithms, here are listed some of the commonly used machine learning algorithms. Each of these algorithms will produce different kind of results; it depends on the size of the data sample used for training and testing the model. Whether the algorithm is supervised or unsupervised it purely depends on the type of data we used. Some mostly used supervised learning algorithms are:

1. Support Vector Machine: SVM is used for classification of data by finding a hyper plane between two classes of data. To find such a plane, SVM tries at maximizing the margin (distance between nearest points) and differentiates between the two classes with small errors. SVM uses a kernel function for classification. This function adds new features to the classes which are linearly separated. SVM gives high accuracy hence; it is widely used to address security problems in IoT and WSNs.

2. Discriminant Analysis- The Discriminant Analysis is a statistical multivariable technique whose purpose is to analyze if there are significant differences between groups of objects regarding a set of variables measured on said objects. [21].

3. Decision Tree- Decision trees are constructed with help of an algorithmic approach that identifies ways to split a data set based on different constrains. It is one of the most widely used and practical method for supervised learning. Decision Trees are used for classification and regression tasks.

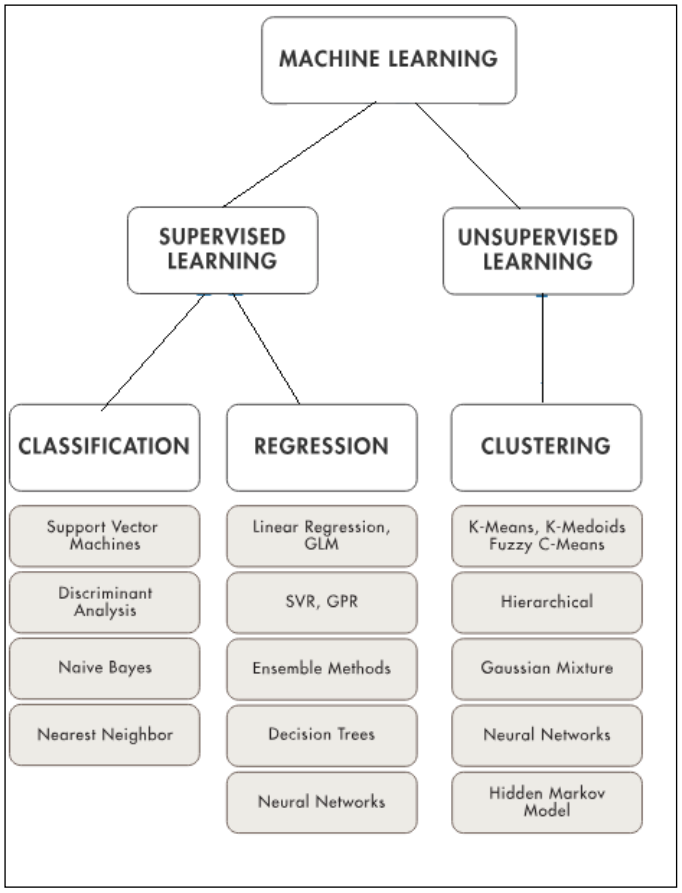

Figure 3- Machine Learning algorithm types [21]

4. Artificial Neural Network- ANNs are models which are motivated by an animal's central nervous systems. These are represented as systems of interconnected "neurons" which can calculate values from inputs. A neural network is an oriented graph. It consists of nodes which in the biological analogy represent neurons, connected by arcs. It corresponds to dendrites and synapses. Every arc associated with a weight while at each node. The values received as input by the node and one can define Activation function along the incoming arcs, adjusted by the weights of the arcs. A neural network is a machine learning algorithm based on the model of a human neuron.

5. K-means Algorithm- K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. The K-means algorithm identifies centroids which are denoted as *k*, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. The 'means' in the K-means refers to averaging of the data; that is, finding the centroid.

6. Naïve Bayes Algorithm- A Naive Bayes Classifier is a supervised machine-learning algorithm that uses the Bayes' Theorem, which assumes that features are statistically independent. The theorem relies on the *naive* assumption that input variables are independent of each other, i.e. there is no chance to know anything about other variables when given an additional variable. It has proven itself to be a classifier with good results.

Table 3: Statistic of WSN model and ML algorithms

IoT Application	WSN IoT protocol	ML Algorithm
Smart Healthcare (Remote Patient Monitoring, Data analytics, appointment reservation)	Bluetooth, ZigBee	Unsupervised learning & supervised learning algorithms (SVM, ANN), Feed forward neural network
Smart Transportation (Intelligent VANET, highway monitoring, speed regulation, driverless vehicle)	Bluetooth, ZigBee	Unsupervised Clustering approach(K means algorithm)
Smart Home (Home security, smart appliances, security system, assistance device)	Insteon, Wavenis, Bluetooth, Wi-Fi, and ZigBee.	Supervised Learning algorithms(SVM, KNN, Naive Bayes) & Unsupervised Learning (k-means)
Smart Building (Energy efficiency, security, smart lightning, attendance monitoring)	Bluetooth, Wi-Fi, z-wave ,ZigBee	Supervised Learning algorithms (ANN, SVM, Bayesian Network) Unsupervised Learning(Hidden Markov)
Smart Education (Digital attendance, tracking, lab safety monitoring, smart access)	Insteon, Wavenis, Bluetooth, Wi-Fi, and ZigBee.	Supervised Learning algorithms(SVM, KNN, Naive Bayes)
Smart Infrastructure (Smart Lightning, automatic weighing, toll management, fault detection)	Bluetooth, Wi-Fi ,ZigBee	Unsupervised Learning Algorithms (K-means algorithm, Clustering and Anomaly detector)
Smart Grid (Energy efficiency, fault detection, monitoring)	ZigBee, 6LoWPAN, RPL, Z-wave, wavenis	Supervised Learning algorithms(SVM, Discriminant Analysis, Decision Trees)

4. CONCLUSIONS

IOT generates massive amount of data with respect to various characteristics and qualities of data. Machine learning combination with IOT gives the persistent development to extend the intelligence of the IOT devices and applications. The exposure of different smart IOT applications with machine learning helps in observation, systematic analysis, processing and smart usages of the large volume of data in different fields. Many industries are using the machine learning and more specifically by using IOT's potential. Hence this paper surveys different WSN IoT protocol and Machine Learning algorithms used with these protocols.

REFERENCES

[1] Jithin Jagannath , Nicholas Polosky , Anu Jagannath , Francesco Restuccia , Tommaso Melodia , Machine learning for wireless communications in the Internet of Things: A comprehensive survey, Ad Hoc Networks 93 (2019) 101913, Elsevier.

[2] K. Ashton , That 'Internet of Things' thing, RFID J. (2009) .

[3] A . Whitmore , A . Agarwal , L. Da Xu , The internet of things –a survey of topics and trends, Inf. Syst. Front. 17 (2) (2015) 261–274 .

[4] Glen Martin (Forbes), How the internet of things is more like the in- dustrial revolution than the digital revolution(https://www.forbes.com/sites/ oreillymedia/2014/02/10/more- 1876- than-1995/#674c4e0b66d2).

[5] L. Da Xu , W. He , S. Li , Internet of things in industries: a survey, IEEE Trans. Ind. Inf. 10 (4) (2014) 2233–2243 .

[6] Ericsson Incorporated, Ericssoninterim mobility report, February 2018, 2018,(https://www.ericsson.com/assets/local/mobility-eport/documents/2018/ emr- interim- feb- 2018.pdf).

[7] Cisco Systems, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper, 2017, (http://tinyurl.com/zzo6766).

[8]Federal Communications Commission [2016], Spectrum crunch(https://www. fcc.gov/general/spectrum-crunch).

[9] H. Shokri-Ghadikolaei , F. Boccardi , C. Fischione , G. Fodor , M. Zorzi , Spectrum sharing in mmwave cellular networks via cell association, coordination, and beamforming, IEEE J. Sel. Areas Commun. 34 (11) (2016) 2902–2917 .

[10] M.A. Vázquez , L. Blanco , A.I. Pérez-Neira , Hybrid analog–digital transmit beamforming for spectrum sharing backhaul networks, IEEE Trans. Signal Pro- cess. 66 (9) (2018) 2273 .

[11] L. Lv , J. Chen , Q. Ni , Z. Ding , H. Jiang , Cognitive non-orthogonal multiple ac- cess with cooperative relaying: a new wireless frontier for 5g spectrum sharing, IEEE Commun. Mag. 56 (4) (2018) 188–195 .

[12] X. Jin , J. Sun , R. Zhang , Y. Zhang , C. Zhang , Specguard: spectrum misuse detec- tion in dynamic spectrum access systems, to appear, IEEE Trans. Mob. Com- put. (2018) ..

[13] Smart health Rayan, Zeina & Alfonse, Marco & M.Salem, Abdel-Badeeh. (2019). Machine Learning Approaches in Smart Health. Procedia Computer Science. 154. 361-368, 8th International Congress of Information and Communication Technology, ICICT 2019. Elsevier Pvt. Ltd.

[14] smart transportation Wireless Sensor Networks to improve Road monitoring by By M. Collotta, G. Pau, V.M. Salerno and G. Scatá, Submitted: December 1st 2011Reviewed: April 19th 2012Published: July 18th 2012.

[15] Smart home Krishna, Sai & Sheshank, Kovi. (2017). WSN'S AND SMART HOME AUTOMATION. 10.13140/RG.2.2.36613.96480.

[16]Machine learning in smart home control systems - Algorithms and new opportunities Ivan A. Berg, Oleg E. Khorev, Arina I. Matvevnina, and Alexey V. Prisjazhnyj Citation: AIP Conference Proceedings 1906, 070007 (2017); Published by the American Institute of Physics

[17] SB- Leveraging Machine Learning and Big Data for Smart Buildings: A Comprehensive Survey Basheer Qolomany, Ala Al-Fuqaha, Ajay Gupta, Safaa Alwajidi, Junaid Qadir, Alvis C. Fong, Senior Member, IEEE

[18]Smar grid -Tuna, Gurkan & Gungor, Vehbi & Gulez, Kayhan. (2013). Wireless Sensor Networks for Smart Grid Applications: A Case Study on Link Reliability and Node Lifetime Evaluations in Power Distribution Systems. International Journal of Distributed Sensor Networks. 2013. 10.1155/2013/796248

[19] International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 21 (2018) pp. 14876-14885 © Research India Publications. http://www.ripublication.com 14876 Machine Learning Techniques and Smart Grid Applications: A Review William Sanchez-Huertas1 , Víctor Gómez1 , Cesar Hernández1

[20] Smart infra- Mohapatra, Badri. (2019). Machine learning applications to smart city. ACCENTS Transactions on Image Processing and Computer Vision. 5. 1-6. 10.19101/TIPCV.2018.412004. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 21 (2018) pp. 14876-14885 © Research India Publications. http://www.ripublication.com

[21] Machine Learning Techniques and Smart Grid Applications: A Review William Sanchez-Huertas1, Víctor Gómez1, Cesar Hernández1*

[22]ANN- https://www.datasciencecentral.com/profiles/blogs/artificial-neural-network-ann-in-machine-learning

[23]Decision tree- https://towardsdatascience.com/decision-tree-in-machine-learning-e380942a4c96

[24]K means- https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

[25] Mohammad Saeid Mahdavinejad, Mohammadreza Rezvan, Mohammadamin Barekatain, Peyman Adibi, Payam Barnaghi, Amit P. Sheth, Machine Learning for Internet of Things Data Analysis: A Survey, Digital Communications and Networks, 10.1016/j.dcan.2017.10.002

[26] Kirti Sharma, Dr. Rainu Nandal, A Literature Study On Machine Learning Fusion With IOT, Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore