Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019)
IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8

A Literature Study On Machine Learning Fusion With IOT

Kirti Sharma[1]
Department of Computer Science and Engineering,
UIET, Maharshi Dayanand University,
Rohtak, Haryana, India
krtbhardwaj1@gmail.com

Dr. Rainu Nandal[2]
Department of Computer Science and Engineering,
UIET, Maharshi Dayanand University,
Rohtak, Haryana, India
rainu_nandal@yahoo.com

***Abstract*— In the present era of rapidly growing mature technologies and their inter-connection with hardware devices and software applications plays a vital role for the emergence of different sensor devices that are inter-connected through internet in order to establish interaction with the physical objects in the world. The technology with the interconnection of different devices through internet is termed as internet of things (IOT). IOT generates massive amount of data with respect to various characteristics and qualities of data. Machine learning fusion with IOT ensures the pervasive 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, the Machine learning-as –a-Service (MLaaS) to exploit the IOT's potential.**

This present paper consists of machine learning basic introduction, machine learning algorithms, reviews of different researcher's study, various sensor devices and the various applications of machine learning algorithms with IOT. And the last section of the paper consists of discussion and conclusion.

Keywords—internet of things; IOT; machine learning; algorithms

I. INTRODUCTION

From the last decades, as world is witness for real time flow of different kinds of structured and unstructured data received from social media, transportation, communication, devices and sensors. As per information, International Data Corporation forecasts that nearby 180 Zeta Byes of data will be generated up to 2025 [1]. This huge explosion of data has given rise to a new economy to the world that is data economy. In today's environment of digital world, data can be considered as the new oil which is pretty important but having worth when it is preceded by cleaned and processed stage [2]. This incremental growth in data economy has given the hike in the usage of concept of IOT or Internet of Things and also incites the existence of data science.

IOT is a hybrid combination between the physical world objects and embedded devices that are connected through the internet to establish the communication. As per day by day involvement of IOT in daily life as well as in business and industry, the more and more amount of data is generated with the help of sensors. Rather than the whole generated newly data it's better to extract the smart data from the IOT data that can be as its best representative [3].

Data Science is also continuously moving ahead in direction to achieve a new paradigm that is Artificial Intelligence which makes possible to teach the machines by learning through the data and propel towards the diversity of its effective insights. Generally, the huge rise in amount of data to exchange makes it impossible to analyze using traditional techniques. So, machine learning intense towards the computation and needs an enormous amount of data for training that includes repetitive training in order to refine the ability of learning as well as decision-taking regarding the applied algorithms [2][4]. Machine learning helps the IOT devices and applications in making the more intelligent decisions with respect to the smart data.

II. MACHINE LEARNING

The discipline of Machine Learning is as a subset of Artificial Intelligence which is concerned with the capability to the computer systems or machines in order to improve the performance automatically throughout its experience [5]. Along with the addition of excessive data, Machine Learning refines the automotive learning process through training and lead towards adaptation of its algorithm.

ML algorithms are used to implement the different models. Basically machine learning can be categorized as: **Guided, Unguided, Semi-guided and reinforcement learning [5].**

A. **Guided Learning**: When learning is applied on labeled data or the desired outcomes is known called as ***guided learning***. 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.

B. **Unguided Learning**: If the learning is applied on unlabeled datasets or the data is not known in advance then this type of learning is known as ***unguided learning [5]***. For example, *NASA* uses this learning approach to create the different clusters of heavenly bodies each of which consists of similar nature objects.

C. ***Semi-guided or Semi-Supervised learning*** approach is a hybrid approach as a combination of guided and unguided learning with few labeled and unlabeled data [5]. For example, *automatically detection of facebook photos* for the multiple photos of a same person from a family function photo cluster (clustering- unguided learning) by naming that person once (guided learning) and afterwards it'll automatically attached the name tag to that person in all photos.

D. ***Reinforcement Learning*** is a machine learning which allows the efficient observation of surroundings and consistent learning behavior to a learning system in order to enhance the frequency of cumulative incentives or rewards [5]. It is also known as *reward-based learning system.* For example, *Robot in a manufacturing unit*.

III. RECENT IOT DEVICES AND TECHNOLOGIES

A. IOT Devices [21]:

- **Arudino Device:** Arudino device is an embedded microcontroller kit used for implementing the digital devices that can sense the objects of the physical world and do the control over them to provide the accessibility and establish communication between those physical objects. Currently, Arudino is one of the popular IOT board used in market that is manufactured with sets of digital and analog I/O Pins that can also be interfaced with other circuits and may include USB Port Connection to connect with PC's.

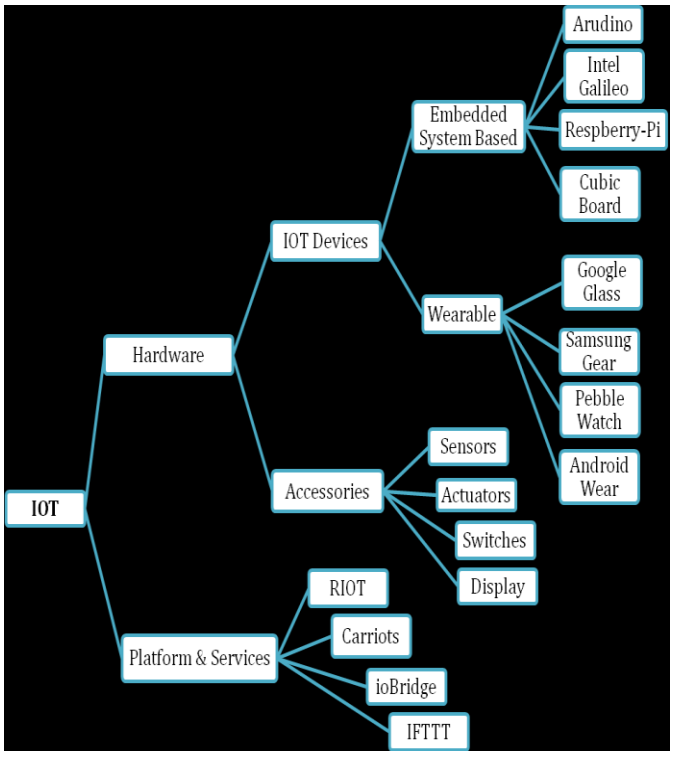

Fig. 1. IOT Devices and Technologies

- **Intel Galileo:** The 2nd generation board of Intel Galileo consist of Intel Quark Super computer processor, 256 MB RAM and different parts in order to provide support to Arudino Devices.
- **Raspberry – Pi:** Most popular IOT board as Pi 3 that provides the more compact and standalone computer systems by involving the Bluetooth, built-in wifi facility and also able to include different programming packages like python, LAMP Stack etc.
- **Samsung Gear Fit:** It is a fitness tracker feature that is present in market as Samsung IOT product. It has curved display, dust as well as water proof and long-lasting battery.

B. IOT PLATFORM

IOT Platform is an integrated service which is used to attain the slot or as an interface between IOT connected devices (Sensors) and IOT applications (Network) and provides offer to bring world's physical objects on-line [22].

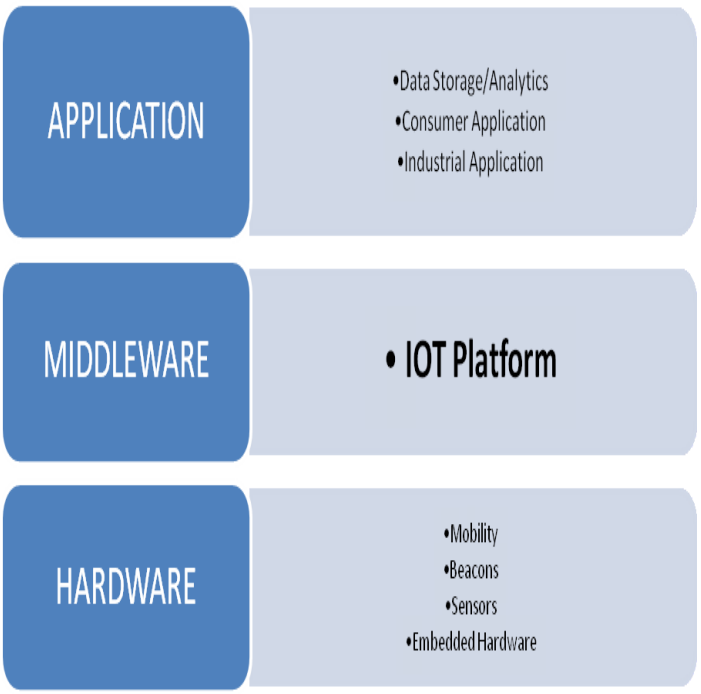

Fig. 2. IOT Platform as Middleware in Architecture

Various IOT Platforms which facilitates to deploy IOT Applications are as:

- **AWS IOT Platform:** Amazon web Services IOT Platform performs the data collection from the connected devices and apply it to the real world scenario by maintaining the security criteria.
- **Microsoft Azure IOT Platform:** Azure IOT Platform provides the bidirectional communication between connected devices and platforms with strong security mechanism, scalability and easy integration with systems. It also included azure streaming analytics which can

process the huge information that is generated by sensors in real world.

- **Google Cloud Platform:** This platform provides some additional features like cloud IOT Core, Speed-Up IOT Device, Cloud Publisher and Subscriber and Cloud Machine Learning Engine. ***IBM Watson IOT Platform and Artik Cloud IOT Platform (SAMSUNG).***

IV. LITERATURE STUDY

IOT is emerging as the challenging area of Information and Technology (IT) and digital world that enlighten the glimpse of further research opportunities in the environment of internet connected devices and applications. Different researchers has been already illustrated their research in this field with some Machine Learning algorithms in order to make proper utilization of huge amount of generated data and to provide some specific point of views.

Lei Zhang and et al. (2012) introduced an approach based on an application for tracking the location i.e. SensTrack which is used with smart phones embedded with Wi-Fi facility in order to reduce the usage of GPS due to its availability at high cost with negative impacts on battery in very short period. SensTrack operated the GPS sample by using stored information and can switch the location and to re-build the track route from recorded location Gaussian Process Regression approach is followed [6].

Mohammed S. Alam and Son T. Vuong (2013) has been used *random forest algorithm of supervised machine learning technique* to classify the Android features and extract the application anomaly as malicious or benign from the given dataset in order to maintain the accuracy during classification. This paper has main focused approach on testing the random forest parameters for the given dataset. During experiment the experimental release of Weka used over the stable release 3.6.9 because newly released version of WEKA provided the multi-threading support in order to minimize the time-complexity level. It also has limitation that WEKA doesn't implement the computation regarding Feature importance in case of random forest classification due to which that's not computed in the paper [7].

Hyo-Sik Ham and et al.(2014) has introduced an technique for the detection of malware in an android platform by using Linear-SVM (Support Vector Machine) on dataset collected by monitoring the resources on android platform to detect the different malwares that were not handled in previous studies and also perform comparative analysis with different labeled classifiers. This paper provides the remedial effect to overcome the drawback of behavior- based approach for detecting the malwares [8].

Divyansh Khanna and et al. (2015) has been presented a comparative study for (multi) classification based methods such that Linear-Support Vector Machine, Logistic Regression and Neural Network for accurate prediction of heart diseases where easier techniques *SVM with linear kernel and logistic regression* provided more accuracy for 5-folds 87.6% and 86.8% respectively. They used the publicly available Cleveland Heart Disease dataset to apply the classification techniques. This study can be used ahead in hospitals to minimize the human-based-error chances in order to contributed to the medical science diagnosis and analysis [9].

Tri Doan and Jugal Kalita (2015) proposed an approach to allot the ranking to the different machine learning algorithms in ranking list with by default setup of parameters by using *linear Regression*. This list of algorithms ranking helped the Auto-WEKA for selection of feasible algorithms for further execution. But with a little drawback that at least four actual datasets features are required that is overcome in their future work in the field of Natural Language Processing and extending the work for big-data challenge [10].

Vishwajeet Hari Bhide and dr Sanjeev Wagh (2015) proposed a novel technique in order to provide the better assimilation of surrounding situations in home with lesser interference of human and also able to automatically detect the fault or problem in any device. Even though, addition of ***Naïve Bayes algorithm*** increased the level of intelligence. Further use case of this system may also able to manage and provide lesser cost factor, efficient amount of energy and flexibility to the smart homes. This paper also included the future scope of their work to include the sms alerts, email notifications or voice alerts with the extended feature to enhance the security at home entrance door by motion detection [11].

Alberto M. C. Soza and Jose` R. Amazonas (2015) has been implemented *as Principal Component Analysis (PCA)* based clustering algorithm for fault detection that used Hadoop Framework and Mahout implementation. This algorithm integrated with IOT architecture implemented by the LinkSmart middleware. Proposed implementation and architecture increased the potential and functionality of IOT LinkSmart middleware [12].

Arijit Ukil and et al. (2016) depicted a ubiquitous health monitoring system based upon (unguided learning) *anomaly detection* of diseases at early stage in order to provide the essence of life. The accuracy of anomaly detection varied with the SNR rate (more accurate with low SNR). This paper also described a use-case of anomaly detection for cardiac with the help of affordable android smart phones with the perspective of early detection of heart related diseases that avoid leading the severe condition [13].

Hardi Desai and et al. (2017) has proposed a vision to implement an affordable and compatible IOT based wireless sensor network in order to monitoring and analyzing the grocery levels at supermarkets as well as at homes. This system also provides an immense to use as future scope in the kitchens and to monitor the different storage places to manage the commodities in smooth manner [14].

Shuai Zhao and et al. (2017) has been proposed a multidimensional model-based approach in order to evaluate the similarities between different IOT services. On the basis of *clustering algorithm* for dynamically discover, match making and exchange will be performed in efficient manner. Evaluation of experiments on proposed system and results are promising and based on real life. It has drawback that in the present paper conducted the practical on single dataset with limited size of test data set. But they also proposed their

drawback to overcome in future work by using the distributed datasets [15].

Guangie Han and et al. (2018) focused on critical issue regarding the security in WSNs to maintain location privacy and introduced the novel protection scheme based upon the k-means that is cluster-based location privacy (KCLP). This particular scheme enhanced the security and minimizes the delay in expense of energy consumption at minor level that can be reduced further in future work [16].

K. Gai, M. Qiu (2018) proposed a novel technique for optimal allocation of resources in cybernetics physical environment (IOT) based upon the reinforcement learning paradigm. Reinforcement method is used to obtain the high satisfactory level that is *QOE (Quality- of Experience),* more accuracy and lesser cost in resource allocation. Reinforcement learning considered QOE level as reward parameter and the resulted value as optimal solution [17].

Srijanee Mookherji and Suresh Sankaranarayanan (2019) has been introduced the IOT Automated Traffic Signaling System in order to dynamically regulate the traffic in congested areas. For data security analysis purpose addressed to Man- In – Middle Attack (MITM) and Support Vector Machine (SVM) based on supervised (guided) learning method is used to deploy at edge to classify the traffic data as raw dataset. This system is implemented on IOT Board Raspberry- Pi3 using Scikit [18].

Jin Ren Ng and et al. (2019) has introduced a motive in perspective to collect the data and with respect to which analyzed conditions of road surfaces in Malaysia using IOT Sensor devices and used supervised (guided) classification algorithm (KNN, Random Forest or SVM). It basically provided the cost effective approach in order to monitor the roads condition which in turn the faster repairing of roads. It helped to provide the better roads that reduce the chances of fatal road accidents due to poor maintenance of roads [19].

Wenjuan Li and et al. (2019) has introduced a combination of design for multi-view based classification of emails for IOT system and the disagreement based semi-supervised learning paradigm. In order to mitigate this major issue of suspicious emails and security a common solution is proposed by taking the two datasets as internal feature dataset and external feature dataset. Basically semi supervised learning paradigm approached both labeled and unlabeled data. The availability of multi-view dataset for email classification provided the more accuracy compared to single view email classification that can be improved further with the help of connecting with different IOT systems [20].

TABLE I. ANALYSIS OF DIFFERENT TYPES OF MACHINE LEARNING ALGORITHMS APPLICATIONS WITH IOT

S No	Year of Work	Objective	ML Type or Model Used	Algorithm Used	Dataset Used	Application
1.	2012	**Energy Efficient Location Tracking with Smart phones for IOT [6]**	Guided Learning-Regression based	Regression Based Gaussian Process	Sensor Generated	Used for Location Tracking with SensTrack. SensTrack followed the location sensing technique based upon Wi-Fi in indoor area.
2.	2013	**Random Forest Classification for Detecting Android Malware [7]**	Guided Learning-classification	Random Forest algorithm	Github Magnum Dataset	For detection of android device to be as the malicious or benign
3.	2014	**Linear SVM Based Android Malware Detection for Reliable IOT Services [8]**	Guided Learning-Classification Based	SVM (linear)	By monitoring the resources in an android environment.	Malware detection which are directly targeted to distinct android platforms.
4.	2015	**Comparative Study of Classification Techniques to Predict the prevalence of heart disease [9]**	Guided Learning-classification based	Support Vector Machine (SVM) with linear kernel/ Logistic Regression	Cleveland Heart Disease Dataset From UCI Repository	To minimize the human-based-error chances in order to contributed to the medical science diagnosis and analysis.
5.	2015	**Selecting Machine Learning Algorithms using Regression model [10]**	Guided Learning-regression based	Linear Regression Algorithm	Random/Unknown dataset	An approach to allot the ranking to the distinct machine learning algorithms in ranking list with by default setup of parameters using linear Regression.
6.	2015	**Automation of an Intelligent Self Learning System for Home using IOT [11]**	Guided Learning-classification	Naïve Bayes Algorithm	Dataset Generated Through Sensors	Provides the better assimilation of environment in home with lesser human interference and automatic fault detection in devices.

7.	**2016**	**IOT Healthcare Analytics as importance of Anomaly Detection [13]**	Unguided Learning	Anomaly Detection	Sensors	In Health Monitoring. To minimize the human-based errors in diagnosis, more reliable system, maximize the early detection rate of diseases.
8.	**2017**	**IOT Service Clustering for Dynamic Service Matchmaking [15]**	Unguided Learning	MDM Algorithm, density-Peak-Based Clustering approach	Single and limited Size Sensors based Dataset	On the basis of clustering algorithm for dynamically discover, match making and exchange will be performed in efficient manner.
9.	**2018**	**Cluster Based Location Privacy in WSN for IOT [16]**	Unguided Learning-Cluster Based	K-means	Sensor Generated	To provide K-means Cluster-based Location Privacy (KCLP) that enhanced the security and minimizes the delay in expense of energy consumption at minor level that can be reduced further in future work.
10.	**2019**	**Traffic Data Classification for Security in IOT– Based Road Signaling System [18]**	Guided Learning-classification	Support Vector Machine (SVM)- at edge	Raw Traffic Dataset generated through 3 cities of London in time period of 5 yrs (2011-2016)	IOT Automated Traffic Signaling System in order to dynamically regulate the traffic in congested areas.
11.	**2019**	**Road Surface Identification [19]**	Guided learning-classification	KNN/ Random Forest/ SVM	Dataset Generated Through Sensors	To provide the better roads that reduces the chances of fatal road accidents due to poor maintenance of roads.

V. DISCUSSION AND CONCLUSION

Machine Learning has an explosive potential to be the key technology for the pervasive environment of IOT. The recent radiations of success of Machine Learning along with IOT prove their fusion as the wonderful source in the field of Data science. This paper discussed the different applications of machine learning algorithms for IOT and covers the recent advanced technologies used in IOT environment. This paper also emphasized the thorough study on recent researches regarding the applications of machine learning algorithms for IOT. In this cutting-edge of artifacts based intelligence of upcoming technologies, the new inventions are reaching at the pinnacle with the perspective of establishing inter-connections in between the physical world objects that provides advanced, smart and reliable life to living being.

REFERENCES

[1] https://www.seagate.com/in/en/our-story/data-age-2025/

[2] Laizhong Cui, Shu Yang and et al. "A survey on application of machine learning for Internet of Things". *International Journal of Machine Learning and Cybernetics*, August 2018, Vol. 9, Issue 8, pp 1399–1417.

[3] Luigi Atzori, Antonio Iera, Giacomo Morabito. "The Internet of Things: A survey". *Computer Networks*, 2010. Vol. 54, Issue 15, pp 2787-2805, ISSN 1389-1286.

[4] Chen, Feng and et al. "Data Mining for the Internet of Things: Literature Review and Challenges." *International Journal of Distributed Sensor Networks*, August 2015.

[5] Mohammad Saeid Mahdavinejad and et al. "Machine learning for internet of things data analysis: a survey". *Digital Communications and Networks*, 2018. Vol. 4, Issue 3, pp 161-175, ISSN 2352-8648.

[6] Zhang, Lei, Jiangchuan Liu, and Hongbo Jiang. "Energy-efficient location tracking with smartphones for IoT". *SENSORS, IEEE,* 2012.

[7] Alam, Mohammed S., and Son T. Vuong. "Random forest classification for detecting android malware." *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*. IEEE, 2013.

[8] Hyo-Sik Ham, Hwan-Hee Kim, Myung-Sup Kim, and Mi-Jung Choi, "Linear SVM-Based Android Malware Detection for Reliable IoT Services," *Journal of Applied Mathematics*, 2014. Article ID 594501.

[9] Khanna, Divyansh, et al. "Comparative study of classification techniques (SVM, logistic regression and neural networks) to predict the prevalence of heart disease." *International Journal of Machine Learning and Computing*. 2015. Vol. 5, Issue 5, pp 414.

[10] Doan, Tri, and Jugal Kalita. "Selecting machine learning algorithms using regression models." *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE, 2015.

[11] Bhide, Vishwajeet Hari, and Sanjeev Wagh. "i-learning IoT: An intelligent self learning system for home automation using IoT." *2015 International Conference on Communications and Signal Processing (ICCSP)*. IEEE, 2015.

[12] Souza, Alberto MC, and Joseé RA Amazonas. "An outlier detect algorithm using big data processing and internet of things architecture." *Procedia Computer Science* 52, 2015. pp 1010-1015.

[13] Ukil, S. Bandyoapdhyay, C. Puri and A. Pal, "IoT Healthcare Analytics: The Importance of Anomaly Detection". *IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, Crans-Montana, 2016. pp 994-997.

[14] Desai, Hardi, et al. "IoT based grocery monitoring system". *Fourteenth International Conference on Wireless and Optical Communications Networks (WOCN)*. IEEE, 2017.

[15] Zhao, Shuai, et al. "Iot service clustering for dynamic service matchmaking." *Sensors, 2017*. Vol. 17, Issue 8, pp 1727.

[16] Han, Guangjie, et al. "KCLP: A k-Means Cluster-Based Location Privacy Protection Scheme in WSNs for IoT." *IEEE Wireless Communications,* 2018. Vol. 25, Issue 6, pp 84-90.

[17] Gai, Keke, and Meikang Qiu. "Optimal resource allocation using reinforcement learning for IoT content-centric services". *Applied Soft Computing*. 2018. Vol. 70, pp 12-21.

[18] Mookherji, Srijanee, and Suresh Sankaranarayanan. "Traffic Data Classification for Security in IoT-Based Road Signaling System". *Soft Computing in Data Analytics*. Springer, Singapore, 2019. pp 589-599.

[19] Ng, Jin Ren, et al. "Identification of Road Surface Conditions using IoT Sensors and Machine Learning." *Computational Science and Technology*, 2019. pp 259-268.

[20] Li, Wenjuan, et al. "Design of multi-view based email classification for IoT systems via semi-supervised learning." *Journal of Network and Computer Applications,* 2019. Vol. 12, Issue 8, pp 56-63.

[21] https://iot.intersog.com/blog/iot-platforms-overview-arduino-raspberry-pi-intel-galileo-and-others/

[22] https://www.i-scoop.eu/internet-of-things-guide/iot-technology-stack-devices-gateways-platforms/