Data Aggregation in Wireless Sensor Networks Using Machine Learning in smart cities

First A. Author, *Fellow, IEEE,*

***Abstract*—Smart cities rely upon wireless sensor network (WSNs) to share data among the smart device. There is a huge number of sensor nodes which are deployed to aggregate data. The sensor nodes are mobile, hence these are charged by batteries which have limited life. While gathering huge amount of data, the power of these nodes goes down, hence it is a major concern to save energy and secure the nodes while gathering data. Another concern is to avoid unnecessary traffic on these networks so that the overall efficiency of these WSNs can be increased. All these aspects of security, efficiency and energy conservation while performing data aggregation can be done using static or dynamic algorithms. The problem with static algorithms is that these are not able to identify malicious data packets and can slow down the network while gathering data from every neighboring node. On the other hand, machine learning can be used to design dynamic algorithms which are self-learning in nature and automatically learn which nodes and destinations to connect to and which ones to avoid. These machine learning methods are a way to achieve high efficiency in WSNs that are actively gathering data. This paper discusses general machine learning methods for data aggregation and some specific tools and methods like tree-based method, grid based, cluster based and neural networks that make data aggregation process more efficient and securer.**

NOMENCLATURE

Artificial Intelligence (AI), Cycle-Based Data Aggregation Scheme (CBDAS), IoT (Internet of Things), Machine Learning (MI), Neural Networks (NN), WSN (wireless sensor networks)

I. INTRODUCTION

THE growing popularity of smart cities encourages scientists to use databases, artificial intelligence, machine learning, networking and sensor wireless related knowledge to make smart cities safer, cheaper and more efficient. Smart cities use remote sensors and Internet of Things (IoT) equipped technologies that gather information from several different locations and help in improving user's quality of life. Smart cities rely upon low power, low data rate wireless sensor networks commonly called WSNs to monitor and control home devices. Nodes lying within WSN make up IoT's core technology infrastructure. These are the embedded physical sensing devices that have internet connection and carry out particular applications. Machine learning and artificial intelligence are playing a great role in automation and data aggregation in wireless sensor networks which might lead to fully automated smart cities in the future. Computer algorithms of machine learning are designed and modeled to improve performance of smart devices based on past experiences, and these experiences are reflected in the form of data available to these devices. The purpose of this paper is to investigate about techniques used for data aggregation in wireless sensor networks and how various techniques differ from each other in improving WSNs. The center of attention are machine learning techniques which help sensor nodes in becoming self-adaptive so that the nodes can distinguish between secure and insecure destinations and decide to accept or reject data packets accordingly. A comprehensive survey based on secondary data

writingsc

shows the direction of scientists at present to achieve highest efficiency of WSNs in the smart cities.

II. Background

Urban buildings are adopting smart cities concepts at a fast pace which has exponentially increased manufacturing and production of IoT devices. In such situation, there is a need to improve the aggregation schemes which can improve performance of sensors in the crowd of networks. Data aggregation, in terms of wireless sensor networks, is a process whereby nodes collect and combine relevant information in a domain which is of interest to them. How effectively nodes communicate with each other depends largely upon which type of information is available to them, and the quality of information is dependent upon data aggregation technique used by these sensors. With the passage of time and advancement in the technology, data aggregation has improved a lot. Besides technical factors, data aggregation also takes into account human factors, thus, artificial intelligence and machine learning have become inseparable parts of data aggregation in WSNs. Certain factors make come WSNs more relevant than others for particular nodes. Machine learning helps sensor nodes identify safer and better WSNs to nodes. Scientists have devised data aggregation models that help in identifying which WSNs will cost less energy to the nodes and which will cost more. Similarly, network capacity can be accurately judged by using the right data aggregation methods. All of these are discussed in this survey paper by critically reviewing existing research.

III. Literature Review

Big data is essential for the performance of smart devices in smart cities, yet it makes it difficult to gather, classify and benefit from relevant data in the IoTs architecture. There are many technologies that help in integration of IoT and big data analytics for smart tourism. Traffic in the wireless sensor networks is increasing which can bring security as well as performance related concerns. By choosing right data aggregation methods, sensors can be directed to decide widely which WSNs to be a part of and which ones to avoid. In order to improve the computing performance of sensor network, data aggregation and traffic planning are done in real-time [1]. Traffic management in smart cities is done using prediction models of machine learning. These models consider payload traffic patterns and periodic traffic patterns within the smart city. In the model, cloud scenario is studied in terms of webserver application in which IoT nodes send the traffic profiles periodically every 30 seconds to gather data about environment. During this process, traffic profiles of sensor nodes are collected at webserver which helps network control attacks on the nodes, thus making network securer [1].

Scientists also have explored four layers model UDP/IP protocol for node connection that resulted into collection of switch layer information and gateway layer information (layer 2 and layer 3 respectively). This approach relies upon 11-byte header length where the said two layers are configured to aggregate as much data as 128 KB that comprises of 16,383 data blocks, each of which is 8-bytes long [1]. This switch aggregation method is used to send data to the IoT gateway that improves efficiency of the network and decreases the number of packets sent, thus resulting into traffic control. Networking specialists look for methods that can use minimum amount of energy for data transfer reducing the noise in the traffic and also decrease energy consumption. Machine learning method of classification is discussed to come over the problems emerging from switch aggregation. The problem with switch aggregation method is that unlike traditional TCP/IP protocol which ensures reliable connectivity, the UDP/IP protocol is computation intensive and includes longer metadata header length. Through k-means method, right switches can be selected.

Another data aggregation method is used for error handling which is called Aggregation Periodic Process (APP) method. This technique analyzes two statistical data aggregation schemes namely constant interval and constant number [2]. The method uses mean total system time to gather data. The efficient networks use space-domain aggregation and time-domain aggregation methods to improve efficiency, but their

writingsc

application is very limited in the smart cities due to long packet length.

Statistical data aggregation methods are proposed to develop adaptive control while access control mechanisms are designed to study how policy can be used to overcome data aggregation challenges in wireless sensor networks. The machine learning based data aggregation models are considered to be most advanced of all the models because these are self-adaptive [2]. Self-adaption of machine learning based sensor networks supports effective decision making whereby malicious data is timely detected and nodes are warned so that they so not accept data from these nodes now or in future. Thus, this systems improves the devices that are part of smart city.

Many WSNs use time-based data aggregation methods. In this method, the source nodes generate data after some time (grooming time of TG); the time required to filter raw data, and then the data is sent to the sink nodes. The method uses time period as the deciding factor about when to share data among nodes. The objective is to update nodes about their neighbors, not to connect only with securer and relevant nodes as shown in figure 1. When the available space is more than data (D<S) the data is continued to get aggregated in the node and when the space is full, data aggregation stops.

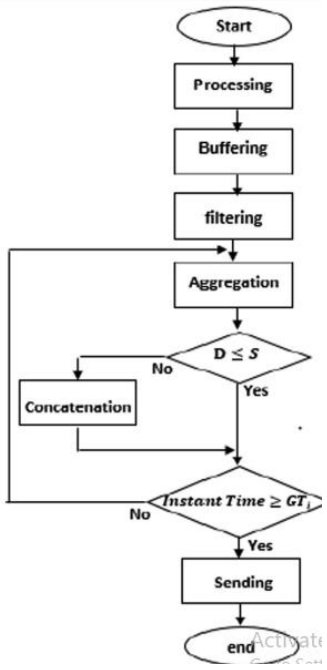
The sink nodes are programmed to gather data bytes originating from multiple nodes and then these nodes forward aggregated data into designed cloud [1]. Using this technique, the WSN reduces number of data packets and headers sent to nodes hence traffic is minimized making smart cities better.

The time-based aggregation method modifies simple time-domain aggregation and optimizes smart city network efficiency by making traffic periodic.

In wireless sensor networks, data is being transferred among the nodes virtually all the time. This data transfer helps nodes in the network in staying updated about their surroundings, about changes in the network and in in identifying any malicious nodes that may affect security of the network. Efficiency is important in this process. In data aggregation, nodes efficiently collect and transfer sensing data to a destination. Machine learning helps sensor nodes in WSNs to methodologically and wisely conduct resource utilization so that data aggregation uses lesser energy and avoids noise i.e. unnecessary traffic.

FIGURE I

Time-based aggregation technique [1]



When nodes open aggregated data that have no noise, their data receiving accuracy and efficiency enhances. Precision is key in efficient data aggregation process which is ensured through machine learning algorithms like neural networks [3]. Machine learning algorithms like neural networks are trained to learn the shortest and securest paths and gather data about network's circumstances and to address the immediate demands of the site where the sensors are situated. Such a high level performance can only be displayed using machine learning algorithms which sense and get rid of errors and noise before aggregating data. Machine learning algorithms use techniques to distinguish trustworthy networks and gateways from untrustworthy ones so that data from trustable destinations is accepted and data from untrustworthy sites and destinations is rejected. These intelligent decisions are made with the help of machine learning algorithms.

Advanced coupling network model (ACNM) is often used to detect problems in packets and to reject them before receiving and opening [3]. When sensor networks use ACNM protocol to deliver packets to networks, the delay in packet transfer is minimised [4]. The delay in packet transfer slows down the overall performance of wireless sensor networks. Fast as well as quick data aggregation is the key to ensuring that the overall network remains efficient as well as noise free. In smart cities; hospital, schools and shops besides homes also have smart devices that are connected to IoT. Healthcare systems in smart cities are concerned about availability, reliability, energy efficiency, security, safety, durability and comfort of smart devices that send and receive data within WSNs. If any of these features is affected and malicious data is found to be sent to destinations, the credibility of WSNs is affected, damaging the trust of patients on hospital systems [5]. It is therefore important that wireless sensor networks are made better in security and data handling and machine learning algorithms are found to do the best job in this regard.

Security and energy constraints concerns are commonly seen in wireless sensor networks within smart cities. Sensor networks are being structurally modified to improve data performance. Some of these structural changes pertain to acquiring proprietary technology for collecting performance data from smart devices, while other changes circulate around using of AI based protocols that improve performance [6]. Low energy consumption algorithms monitor battery powered nodes so that it can he ensured that nodes are not spending energy on data aggregation from incorrect or blocked/malicious gateways or networks. Data aggregation in wireless sensor networks is complex than any other network because of the size and scope of these networks and the number of nodes adding to and leaving these networks. Manual handling of data question is out of question in these networks, and static decision making (to pick and choose packets) is also impractical for such large networks. Machine learning based dynamic methods help, on one hand to aggregate data, and on the other hand, these have self-learning capacity which means these networks improve with time because of ML implementation.

Energy efficiency issue in WSNs is also discussed in [7] where the research explores distributed sensors in WSN. These sensors are present in different parts of smart cities and the environment of devices varies from time to time. During data aggregation in these networks, data redundancy and accuracy issues are faced. To resolve these issues, different machine learning techniques are used particularly a three-layer neural network is used for efficient backward propagation. Performance of network improves when input layer neurons, which are situated at members of each cluster as the hidden layer neurons lie in cluster heads, communicate with output layer neurons that are found in base station [7]. When the three-layer NNs are used in simulation, it shows that data aggregation becomes accurate as well as efficient and the network consumes lessor amount of energy. Nodes within Neural networks act as smart neurons so they receive and process data by linking nodes and processing information using fuzzy logic.

Another machine learning technique that improves data aggregation in WSNs within smart cities is clustering technique. Clustering analysis is unsupervised machine learning method whereby the nodes automatically discover natural grouping within data. In the contrary approach of supervised learning, the nodes adopt predictive model and when given input, nodes can predict output. In clustering, it is not known at initial stages that what will be the mean or central criteria of making clusters [8].

Wireless sensor networks can be of different sizes. These can be small networks within a campus or can compromise whole cities. The range of wireless sensor networks varies upon range of devices that are part of them. The major limitation of the wireless sensory networks is that communication range is about 100 meters, but an advantage of this range is that it helps in maintaining security as well[8]. Using this technique helps nodes choose which neighboring nodes can be connected with for data aggregation purpose, and which nodes are irrelevant, hence data should not be sent to or received from those nodes.

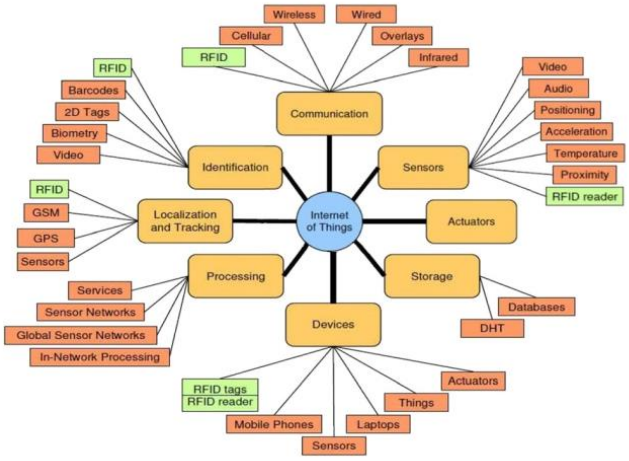
There is a huge number of tools involved in IoT as shown in figure 2. Some of the commonly used devices are laptops,

Writings

mobile phones, sensors in home appliances, actuators and RFID tags. Storage devices that are part of IoT include, but are not limited to databases, and the sensors include video, audio, acceleration and temperature sensors, among many others. Communication things in IoT include RFIDs, wireless, wired and infrared communication devices as shown in figure 2. Similarly, there are identification and localization things which are also major part of IoT. These send and receive, millions of packets per day and it is a huge task to manage data emitting and coming to these sensory nodes. The tools are ever increasing in number which makes their integration important as well as complicated [8].

FIGURE 2

Tools involves in IoT [8]



Smart objects are increasing in number and diversity and there is a dire need for collecting and analyzing data generated by these devices so that these can be well integrated into the smart homes, which in-turn make up the smart cities. Smart devices or the sensor nodes are battery-powered which increase the device mobility on one hand, and restrict efficient operations on the other because energy has to be carefully utilized. By sensing and aggregating irrelevant data, these sensory nodes can consume battery power too early, making the system inefficient. This problem requires sensor nodes to eliminate redundancies which exist in data packets received from adjacent nodes [8]. By choosing right data aggregation methods, sensory nodes can avoid collecting redundant data and also these can reduce network traffic.

Complexity and scope of smart cities is increasing by every passing day. More and more smart devices are entering the networks, some of which are computing devices while the others are home appliances, security devices, alarms and alike. The huge amount of packets being received from all neighbors make it difficult for a sensory node to pick which packet is to be accepted, and which should be rejected because of malicious or irrelevant information within. In order to streamline the bulk volume broadcasted and to organize inconsistent and heterogeneous devices, multiple hierarchal approaches are used to sense data packets and extract information from these. The cluster method is one of these where sensors use information from cluster heads to judge whether data in packets is relevant to them or not. The cluster head node for the cluster is responsible for routing [8]. Chain-based or string based methods for data aggregation are considered energy efficient where energy distribution takes place equally at each node and these nodes are associated with adjacent nodes. The nodes exchange TOKEN with the leader node and upon receiving TOKEN, the sensory node sends information to the node that performs data aggregation from where data is directed towards Sink station [8]. This resolves the issue of traffic overload by eliminating immense overhead of real-time dynamic cluster.

Figure 3 shows some selected data aggregation tools of machine learning used in WSNs. Another popular data aggregation methodology in machine learning is tree based data aggregation. The tree-method aggregates data by creating tree of data aggregated from nodes. Spanning tree can be made which is later optimized to reduce the number of transmission, resulting in prolonged lifetime of the sensor network [8].

Comparing tree-based method to the cluster-based method, it can be found that the later results into average overhead, a low energy uniformity, average strength and flexibility, scalability and low power consumption. On the other hand, the tree-based methods cause high overhead and energy uniformity, strength, flexibility, scalability and offer average energy consumption.

Writings

FIGURE 3

Data aggregation based on network architecture [8]

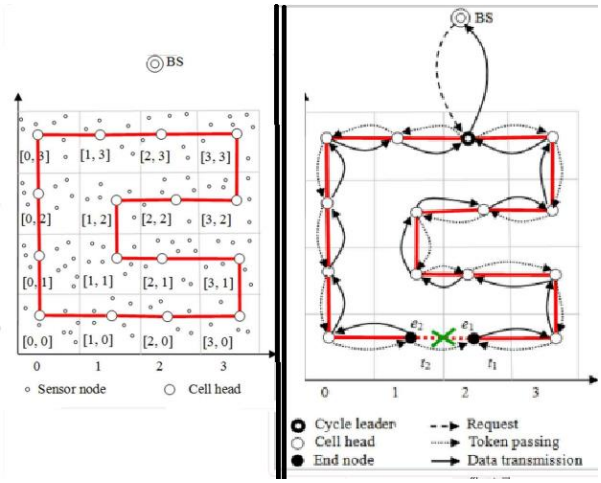
These two methods are superseded by the grid-based data aggregation which comprises of the array of sensors which are integrated in constant region of the sensor network. The nodes send data to grid directly which frees nodes from the requirement to communicate with each other [9]. The grid-based data aggregation resembles with cluster-based data aggregation method such that in both these methods cluster head is fixed. Military surveillance networks use secure grid-based data aggregation.

Another important data aggregation machine learning method is Cycle-Based Data Aggregation Scheme (CBDAS) which is used in grid-based WSNs [9]. Sensor field with in CBDAS is divided into a grid comprising of cells which have heads. By connecting these cells via heads in a chain-format, the WSN prolongs lifetime of the network. Data gathering takes place in rounds where data moves from one node to another. Data transmission only takes place through cell heads in cyclic chain which slows down energy depletion, allowing battery life to survive longer.

Unlike the cluster-based protocol, in which nodes group themselves based on similarities and gather data from similar nodes, the grid-based protocol establishes a grid on a per-source basis [9]. Figure 4 compares cycle based data aggregation to grid based data aggregation.

FIGURE 4

Tools involves in IoT [8]



Data is transmitted by a source and it is carried out after constructing a grid. The points lying within the grid have dissemination nodes which perform the task of data storage and data transfer [9]. The role of the sink is to transfer information as soon as it receives it. Through message communication without a delay, the high speed of network is ensured. When messages are forwarded to the sink, a connection between source and sink is created. Grid has many coordinators that establish many paths and make it easier for future messages to be sent via those paths. Grids make message-communication faster in the networks.

When obtaining data from various sensory nodes in a network through process of data aggregation, the objective is to minimize redundant data so that the traffic is controlled and duplicated or similar messages are not sent multiple times. Several techniques in machine learning, as discussed above, help in selecting data carefully and minimizing it accurately [10]. Figure 5 shows different machine learning techniques that can be used to find patterns in data that can help in identifying relevant information and eliminating data that is not of interest to the network. There are supervised, unsupervised, semi-supervised and reinforced learning techniques within machine learning, many of which are used to decide which information to keep and which to drop while performing data aggregation.

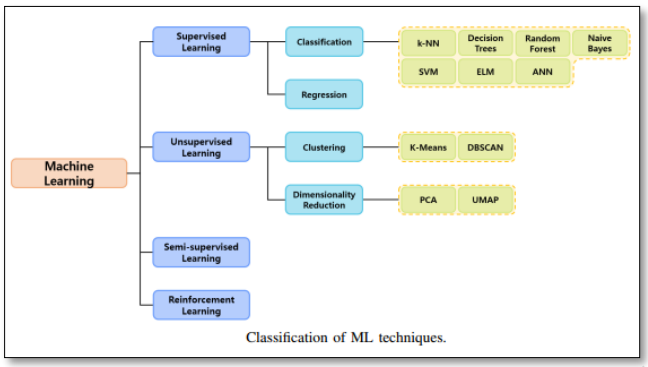
writings

Besides controlling data redundancy while performing data aggregation. Scientists try to achieve efficiency within wireless sensory networks by making battery power last longer. This is done by minimizing energy consumption without compromising on the performance of networks. Sensory nodes cannot be shut down for security and productivity reasons, and hence the solution is to use intelligent methods for gathering data, rather than performing data aggregation continuously. Besides time-based data aggregation, the clustering and tree based methods are used to decide when to aggregate data, but as the complexity of networks increase, more detailed methods of machine learning are required to perform the job. Energy-focused data clustering and data reduction uses machine learning classifiers such as decision trees, support vector machines and neural networks to identify and classify similar data [10]. With the help of ML methods and algorithms, WSMs direct nodes to use attribute-value pairs whereby node requests data through sharing interest messages for named data. Upon receiving these requests, neighboring nodes become active and help source node in drawing relevant information towards it. Data from the cache memory of the nodes is used through directed diffusion which is a technique that helps flood low-rate data to establish a reinforced path between destination node and the source node.

Once the flooding is caused initially and data from neighboring nodes starts pouring towards the source node, the node decides which nodes brought the relevant data and then records those nodes, creating a path. In the future, when a relevant information is required, the source node does not flood data and rather uses the established path to gather data. This helps in reducing the amount of energy required to operate (energy conservation and long life of the network are guaranteed) and also helps in making a secure and reliable connection that is least likely to allow malicious data from external sources. These paths are established using NN or Decision tree which are most commonly used machine learning methods [10]. Besides data aggregation in the normal routine, there are cases when certain nodes have lower energy and these cannot often help in aggregating data. In such cases, the alternate routes can be established to ensure connectivity and to perform data aggregation continuously.

FIGURE5

Machine learning methods and subtypes [10]



Decision making is not only important, but also difficult in wireless sensor networks where nodes are entering and leaving the network all the time. The size of the networks and the type of nodes within it help in deciding which type of tools are better suitable for them to aggregate data. Complex systems and nodes are more likely to benefit from neural networks while the simpler networks can use support vector machines or even decision trees to decide how to select path when sending message from a source to a destination [2] [10].

writingsc

TABLE 1
MACHINE LEARNING METHODS AND SUBTYPES [10]

Sr. No	Year	Authors	Journal	Techniques used	Preferred Technique	Disadvantages	Advantages
[1]	2022	Ibrahim et al.	Alexandria Engineering Journal	Classification	k-means	Lengthy Overhead is high	Efficient
[2]	2021	Ibrahim et al.	Adv. Sci. Technol. Eng. Syst. J.	Classification, Clustering	Clustering	Sometimes output is not given, so classification is useless in such cases	Clustering is efficient when the output is not known
[3]	2022	Sudha et al.	Theoretical Computer Science	Classification methods	Neural Networks	Laborious to deploy	Better error handling, Noise reduction
[4]	2018	Mohamma et al.	International Journal of Engineering and Technology	Clustering method, sleep/wake scheduling	sleep/wake schedulin g	Energy loss when aggregating data perpetually	Energy efficient clustering is performed using sleep/wake scheduling
[5]	2013	Hadjidj et al.	Journal of Networkand Computer Applications (JNCA),	Classification (gait analysis, activity recognition)	Classificat ion using K-means for gait analysis	Storage problems	enable real time configuration of sensor sampling rat
[8]	2019	Dehkordi et al.	Iireless Networks	Clustering	Grid-based clusters	High traffic in absence of clustering technique	Broadcasting is avoided and few transmission resolve communicatio n requirements
[10]	2020	Kim et al.	IEEE Explore	Decision tree, random forest, naïve Bayse , K-means, NN	All methods have certain merits	DT help in simple and static scenarios	Diverse applications help in different situations

Various tools and techniques of machine learning are used in different researches, as discussed above, but the goals of these researches are the same, which are, to minimize energy consumption, to increase lifetime of the networks and to enhance the accuracy of networks, as shown in the figure 6. Accuracy issue of wireless sensor networks is also discussed in [11] where the focus is on the accuracy of data collected as well as accuracy of information about whereabouts of the nodes that make up the reliable path to the destination. Paper suggests that accurate and sufficient location information is the mandatory for maintaining wireless sensor networks applications. It also discusses the problem of incomplete and misleading computations of network ranges when dealing with current range-based localization approaches. Better and more advanced decision making approaches are suggested so that noise and measurement issues can be handled in networks [11]. The paper also takes into account the fact that certain noises are unavoidable in the WSNs and these are considered as structural

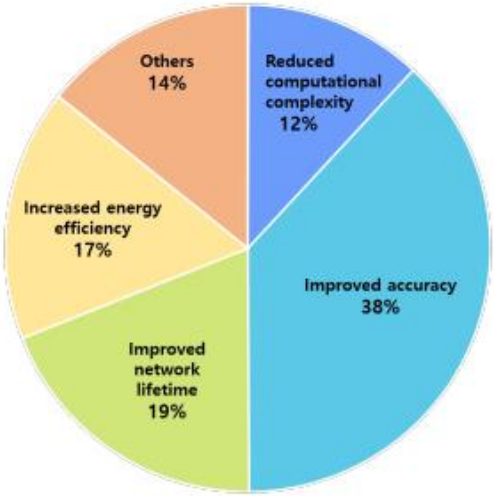
Writings

noises, which are hard to be removed from networks without makings structural changes in the architecture of WSNs.

The smart cities have numerous sub-systems which rely on IoT. From traffic management systems, to e-health and education and e-agriculture, every subsystem has hundreds or even thousands of devices in order to maintain their efficiency, machine learning tools are used to perform automated operation of the internet of things (IoT) nodes. The type of information generated and needed by traffic monitoring system, smart waste management, smart buildings and e-healthcare varies yet these are accessed by common administration which needs to aggregate data in universal format so that data access, data retrieval and decision management based on this data is convenient.

FIGURE 6

Tools involves in IoT [10]



Impact of using ML-based algorithms on performances of WSNs.

Machine learning techniques are used for data testing and data training so that the smart cities are enabled for efficient decision making [12]. Paper proposes that machine learning methods optimize regular WSN-IoT nodes and help in accurate data aggregation.

The problem with machine learnings are also discussed in [12]. The paper argues that although decision making is ultimately streamlined using technology, yet it takes a lot of time to train data. Huge data sets are required first to prepare the software so that it gets familiar with the system variables and data, only after which it is possible to use system for data testing and decision making. Decision tree method is preferred in the paper to train WSN nodes so that the smart city can accurately handle its data.

In a research, challenges facing IoT-enabled smart cities are discussed describing how IoT technologies contribute to the Smart City and how these can providing better and securer lifestyles to the residents of smart cities. Some of the techniques of machine learning are discussed in the paper that improve functionalities of smart cities. These techniques are supervised, unsupervised and reinforce learning methods [13].

The supervised data is used in smart cities learn how to map a set of input values to given values of output. This cannot be done in data that is not labelled. When the devices generate data regarding environment for instance, the smart cities can use this data to predict weather in terms of hot, cold, rainy, sunny, cloudy etc. However, data aggregation in some areas of traffic management will involve parts of supervised learning and parts of reinforced learning where decisions are made on the basis of real time data as it is generated. Data generation and decision making in real time is difficult for nodes and these may utilize reinforced learning for better decision making.

Research shows that machine learning algorithms in WSN-IoT open up performance doors for smart cities, yet bring many challenges as well [14]. The accuracy of different machine learning methods is different for accurately aggregating dat. It was found that supervised learning algorithms guarantee around 61% as accurate results while data aggregation in regression learning is no more than 27% and while unsupervised offers only 12% accurate results [14]. However, this accuracy may vary from case to case. The smart cities should select suitable algorithms that align with their problems [15]. A study shows that the network scientists will need more than one supervised

learning methods to perform accurate data aggregation in different domains of IoT-enabled smart cities [15]. It is also widely accepted that the distributed nature of databases is an additional challenge for smart cities which makes it difficult for the mobile agents to aggregate data in wireless sensor networks [16].

The distributed databases are also need of the hour which is why wireless systems rely on them. The solution to this problem is to not aggregate bulk of data and then transfer it to sink nodes, rather sensor nodes are facilitated to collect data [16]. Machine learning allows nodes to decide when data has to be collected and transferred and when to stop collecting further data as the storage nears full capacity. This increases fault tolerance of the WSNs in smart cities and makes them efficient. Many mistakes can be potentially made when sensing the environment, collect the data, and transmitting it to the base station within using wireless sensor network. The networks have to avoid faults as well as have to conserve energy [17]. Synchronized tree and decision tree approaches are commonly deployed to aggregate data efficiently. A research used fuzzy attribute-based joint integrated scheduling and tree formation (FAJIT) method to form trees and select nodes that are best suitable for data transmission [17]. The method helped heterogeneous networks develop reliable connectivity for communicating messages among nodes.

Local collision can often take place between the nodes which can be avoided using scheduling so that the nodes can continue to aggregate data as per their suitability [18]. The computation transmission schedule helps every node to have a bottom-up manner level schedule that shows time to send data to the parent node. The machine learning tools are used to update schedule and share it with subset of its neighbors through broadcasting as discussed in [8]. After broadcasting, the paths are chosen which mean in the future, the selected paths are used for data transfer, hence no need for broadcasting. It is found that during the planning stage of smart cities, the network topologies and intelligent systems should be decided which will help in the integration of appliances and devices in IoT, because if the matter is ignored during planning stage, it will be far challenging at later phases of smart city implementation [19]. When the smart city administration does not choose between grid or cluster based protocols in advance, the full potential smart cities cannot be enjoyed [20].

IV. Conclusion

An abundant amount of research is available on the data aggregation in wireless sensor networks using machine learning methods in smart cities. Network scientists have explored how supervised, unsupervised and reinforced learning methods can be used to improve the decision making capacity of WSNs so that data aggregation can be carried out in efficient and accurate manner. It is found that majority of researchers agree to using supervised learning methods like decision trees and neural networks to enable smart cities for better data aggregation. Yet, the accuracy of data aggregation by nodes varies from scenario to scenario. This calls for careful choice of ML methods based on circumstances under consideration. By making a wrong choice, the system efficiency can be compromised which can have serious consequences like inaccurate data aggregation, short life of wireless networks and quick battery discharging issues. To safe energy and to avoid redundant data, right machine learning algorithms must be chosen. Growth of smart cities demands these most advanced methods to be used in place of static methods.

V. References

[1] A. S. Ibrahim, K. Y. Youssef, Eldeeb, A. H., M. Abouelatta, and Kamel, H. (2022). Adaptive aggregation based IoT traffic patterns for optimizing smart city network performance. *Alexandria Engineering Journal.* [Online]. 61. 9553–9568

[2] A.S. Ibrahim, K.Y. Youssef, M. Abouelatta. (2021). Traffic aggregation techniques for optimizing IoT networks, *Adv. Sci. Technol. Eng. Syst. J.* 6 (1). 509–518

[3] C. Sudha, D. Suresh and A. Nagesh. (2022). Accurate data aggregation created by neural network and data classification processed through machine learning in wireless sensor networks. *Theoretical Computer Science,* Volume 925, pp. 25-36

[4] H. Mohammad and A. S. C. Sastry. (2018). ACNM: advance coupling network model sleep/awake mechanism for wireless sensor networks. *International Journal of Engineering and Technology*. Vol. 7. No. 1

writingsc

[5] A. Hadjidj, M. Souil, A. Bouabdallah, Y. Challal, H. Owen. (2013). WirelessSensor Networks for Rehabilitation Applications: Challenges and Opportunities. *Journal of Networkand Computer Applications (JNCA), Elsevier*, vol. 36, pp.1-15.

[6] G. Sharma, M. Verma, N. Mishra. (2014). Analysis of transmission technologies in wireless sensor networks, *International Journal of Engineering Research & Technology (IJERT),* Vol. 3 Issue 1.

[7] F. Khorasani and R. Naji, (2017). Energy efficient data aggregation in wireless sensor networks using neural networks. *International Journal of Sensor Networks.* Vol. 24. No. 1.

[8] S. A. Dehkordi, K. Farajzadeh, K. Rezazadeh, Farahbakhsh, R., K. Sandrasegaran and M. A. Dehkordi, M. A. (2019). A survey on data aggregation techniques in IoT sensor networks. *Wireless Networks.* Vol. 26. Pp. 1243–1263

[9] Y. K. Chiang, N. C. Wang, and C. H. Hsieh, (2014). A cycle-based data aggregation scheme for grid-based wireless sensor networks. *Sensors (Basel),* Vol. 14. No. 5, pp. 8447–8464.

[10] T. Kim, L. F. Vecchietti, K. Choi and S. Lee. (2020). Machine Learning for Advanced Wireless Sensor Networks: A Review. *IEEE Explore.*

[11] F. Xiao, W. Liu, Z. Li, L. Chen, and R. Wang, (2017). Noise-tolerant wireless sensor networks localization via multinorms regularized matrix completion, *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2409–2419.

[12] H. Sharma, A. Haque and F. Blaabjerg. (2021). Machine Learning in Wireless Sensor Networks for SmartCities: A Survey. *Electronics.* Vol. 10. Pp.1-22

[13] Bauer, M., Sanchez, L., and Song, J. (2021). IoT-Enabled Smart Cities: Evolution and Outlook. *Sensors*. Vol. 21, 4511.

[14] H. Sharma, A. Haque and F. Blaabjerg. (2020). Machine Learning in Wireless Sensor Networks for Smart Cities: A Survey. Electronics. vol. 10. iss. 9.

[15] M. A. Alsheikh, S. Lin, S. and Niyato. (2014). Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Commun. Surv. Tutor*. Vol. 16. 1996–2018.

[16] M. E. Fissaoui, A. B. Hssane and M. Saadi. (2019). Energy efficient and fault tolerant distributed algorithm for data aggregation in wireless sensor networks. *Journal of Ambient Intelligence and Humanized Computing*. Vol. 10. Pp. 569–57

[17] S. Bhushan, M. Kumar, P. Kumar, and T. Stephan. (2020). FAJIT: a fuzzy-based data aggregation technique for energy efficiency in wireless sensor network. *Complex & Intelligent Systems*. Vol. 7. pp.997–1007

[18] Q. Chen, H. Gao and Z. Cai. (2018). Distributed Low-Latency Data Aggregation for Duty-Cycle Wireless Sensor Networks. *IEEE/ACM Transactions On Networking*

[19] Du, R.; Santi, P.; Xiao, M.; Vasilakos, A.V.; Fischione, C. (2019). The sensable city: A survey on the deployment and management for smart city monitoring. *IEEE Commun. Surv. Tutor*. 2019, 21, 1533–1560.

[20] Morello, R.; Mukhopadhyay, S.C.; Liu, Z.; Slomovitz, D.; Samantaray, S.R. (2017). Advances on sensing technologies for smart cities and power grids: A review. *IEEE Sens. J*, 17, 7596–7610.

writingso