Implementation of Fusion and Filtering Techniques in IoT Data Processing: a case study of Smart Healthcare

Sultan Ahmad[1]*, Abdullah Alharbi[2], Abu Sarwar Zamani[3] and **Mohammed Yousuf Uddin[4]**

[1]Department of Computer Science, College of Computer Engineering and Sciences,
Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia
[2]Department of Information Technology, College of Computers and Information Technology,
Taif University, P.O. Box 11099, Taif, 21944, Saudi Arabia
[3]Department of Computer Science, Deanship of Preparatory Year,
Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia
[4]Department of Information Systems, College of Computer Engineering and Sciences,
Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia
*Corresponding Author: Sultan Ahmad

Summary
Huge data volumes are generated by connected devices, every organization try to make use of its full potential. IoT data is more valuable than ever but without analytic power, data generated by IoT devices and solutions would offer limited added value and insights. We have explored the role of proper data processing and analytics in IoT and find out what are the key steps to maximizing the value of data generated from IoT devices and also discussed the fundamental steps required to turn any organization's IoT data into valuable information for the business. Key questions should be answered by any organization that is embracing and offering IoT solutions and products, prior to embarking into IoT data analytics projects, to ensure success and avoid common pain-points. This paper will provide a structured approach using fusion and filtration techniques to truly transform IoT data into actionable business information and create a long-term and future-proof IoT data strategy.

Key words:
Fusion, Filtration, Fog Computing, IoT Data, Edge Computing, IoT, Cloud Computing.

1. Introduction

In recent times, the numbers of smart devices are more than the number of people living in the world. Among those, maximum people are connected throughout the day. The numbers of devices are increasing per person day by day. For example, now a day's people are not relying on only one device rather they use more than one. These devices include smart phones, tablets, health and exercise monitors, tablets, kindle and so on. As per the prediction the number of devices per person will increase in 2020 by an average of 6.58. As per the prediction of IDC, the connected devices will reach 75 Billion by 2025[1]. Three by Fourth of these will be related to IoT. IDC also predicted that the volume of data will be 73.1 Zetta Bytes by 2025 [2] that was 18.3 Zetta Bytes in 2019 as shown in the figure 1. These data mainly belong to surveillance and security sector. Industrial IoT will also have a good proportion. All of this is possible because of the modern day facility of connectivity. The connectivity of the world is covered quickly from region to region that permits the smart devices to transmit and interconnect with the world. The network connectivity is like a mess or one can assume that it is a layer like ozone surrounding the earth. The digital layer allows the devices to connect and transmit the message from one to another. The layer also allows communicating from one to another, monitor the movements digitally, evaluate if necessary, and in many situations transmit and collects the data and adjust it.

As a result, the digital devices are being embraced by the society; it continued to accept by the globe. As a consequence, it accelerates the benefits in terms of economy and promotes the digital transformation. The application of digital transformation provides platform for innovation in business and industry. This digital transformation and technical advancement used by the human society to enrich the communication and reach of individual to the world.

The Internet of Things (IoT) is the connection of millions of smart devices and sensors connected to the Internet. These connected devices and sensors collect and share data for use and evaluation by many organizations. These organizations include businesses, cities, governments, hospitals and individuals. The IoT has been possible, in part, due to the advent of cheap processors and wireless networks. Previously inanimate objects such as doorknobs or light bulbs can now be equipped with an intelligent sensor that can collect and transfer data to a network.

Researchers has predicted that more than 3 million new devices are connected every month to the Internet.

Manuscript received December 5, 2020
Manuscript revised December 20, 2020
https://doi.org/10.22937/IJCSNS.2020.20.12.14

Researchers also estimate that in the next four years, there are going to be over 30 billion connected devices worldwide. Computers, tablets, smart TVs, and smartphones are one third of the amount of connected devices available for now. Two- thirds which is remaining will be another kind of things like sensors, newly invented intelligent devices, actuators that analyses, optimize and control our world.

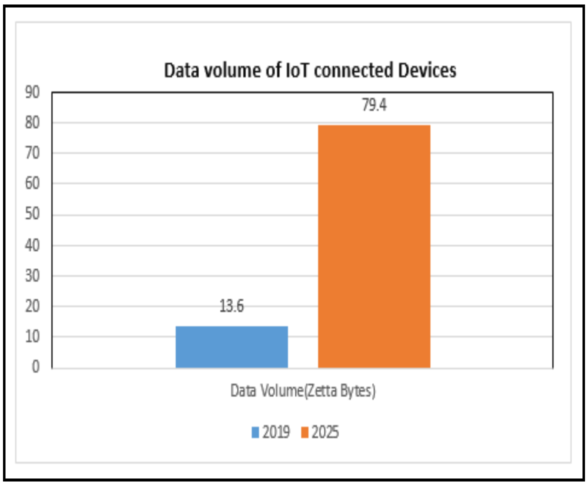

Fig. 1 Data volume of IoT connected devices worldwide 2019 and 2025 (Source - Statista ,2019)

Some examples of intelligent connected sensors are: smart doorbells, garage doors, thermostats, sports wearables, pacemakers, traffic lights, parking spots, and many others. The limit of different objects that could become intelligent sensors is limited only by our imagination. As a result, computation resources and storage space are required for these bulk of data generated by these devices.

2. Problem Statement

As the huge generated data by IoT devices are required proper handling for their compute, store and retrieve for further use. Moreover, the requirement of all these computation storages cannot happen at the network edge. Consequently, cloud computing is the only way which provides solution to storage of massive data by giving infrastructure through powerful applications. However, the problem with cloud computing is that it does more suitable for real-time response and also encounter problems with mission-critical applications with strict runtime and latency requirements [3]. Additionally, billions of IoT devices needs processing, storage, and communication demands, which it cannot scale sufficiently handle. To get rid of this problem Fog computing offer you additional storage, control capabilities that provides additional commuting closer to the network edge. Fog Computing can support IoT data management by using appropriate classification, Fusion and Filtration methods and algorithms. These techniques may be referred to Smart Healthcare ecosystem that is also effected with digital transformation and IoT technologies.

3. Related Works

Although researchers continue to tackle IoT devices issues of security and privacy, huge data generated by these growing technologies, need to be addressed more. Data Fusion technique is one of the effective way to handle these great volume of sensor generated data.

Callegaro et al. (2014) proposed an architecture for the fusion of information aimed at low cost sensors in Wireless Sensor Networks(WSN) [4]. They divided the architecture into three layers: local fusion (occurs in the sensor that collects the data), low level fusion (occurs at the network coordinating sensor or later offline) and management (user interface and medium and high level mergers). The authors limited to specify the operation only the low - level layer, focusing on the analysis of detection algorithms outliers for increased accuracy of data.

Wichit (2014) has a multisensory data fusion architecture with a focus on human behavior recognition using fuzzy logic-based fusion algorithm (fuzzy) [5]. The authors do not specify a layered structure for architecture, but a flow for the recognition of human activity and the application of inference with diffuse logic.

On the same estimation, Wang et al. (2015) suggested two-tier structure of data fusion focusing on improvement of data accuracy [6]. The first tier level, sensors are available which merge and group the same data structure. The second tier level perceives the grouped and merged data structure of the first tier and performs its fusion using the algorithm of covariance intersection. The authors create rules for increasing the accuracy of the data without focusing on the interpretation of the data.

De Paola et al. (2016) propose a “context-sensitive data fusion system, self-optimized and adaptive, based on a three-tier architecture” [7]. The lower layer is responsible for sensing, generating raw data. The middle layer fuses data, in which it tries to integrate information available in the context. And the upper layer seeks to achieve a balance between system performance and execution costs (such as energy consumption). But it was not specified whether any of the layers performs merging decisions or what the interface for creating rules and management would look like. As a proof of concept, use the activity detection scenario in an intelligent environment.

Bish et al. (2016) describe a “fusion architecture that combines multimodal and multisensory fusion within the framework of the Open Standard for Unattended Sensors - Open Standard for Unaccompanied Sensors (OSUS)”. The goal of the authors' architecture [8] is to be a modular, plug and play system, and that fusion methods that may exist in the future are easily integrated. This proposal focuses on how the structure proposed by the authors works in the OSUS framework, where multisensory fusion is one of the layers of the model, without going into details about the functioning of this layer.

Segall, R et al. (2018), has provided conception of vast data in cloud and fog computing in IoT environment [9]. The authors have conducted the case study for actual data from fatality analysis reporting system (FARS) controlled by the National High-way Traffic Safety Administration (NHTSA) of the United States Department of Transportation (USDoT). The study explains the challenges and scope of prospects in using Big data with Fog computing.

Therefore, more works are required in order to handle the vast increasing big data. It is necessary to have safety measure and related effects of these huge data [10]. Here, we have taken the smart Healthcare for our case study and experiment to explain the fusion and filtering methods in automation and proper handling of healthcare data in IoT environment. We explored how can fusion and filtering can be used to handle data of smart healthcare system in order to make more productive and easier handling of patients and hospital requirements.

4. Characteristics of IoT Data

IoT resources are continuously generating data. The massive amounts of data streams have particular characteristics that are different from conventional data streams. The data is collected from heterogeneous resources with a different format and various quality and granularities. The data is not only large in volume, but it is also continuous, dynamic with spatial and temporal dependency. IoT data streams are often collected and published with meta-data, and consequently, the streams have a wide variety of representations.

IoT data is a type of big data. There are five intrinsic characteristics of big data (5V's) [11]; Volume, Variety, Velocity, Veracity and Value. The variety is increasing while technology is advancing and the amount of data is growing while network-enabled devices are connecting to the Internet.

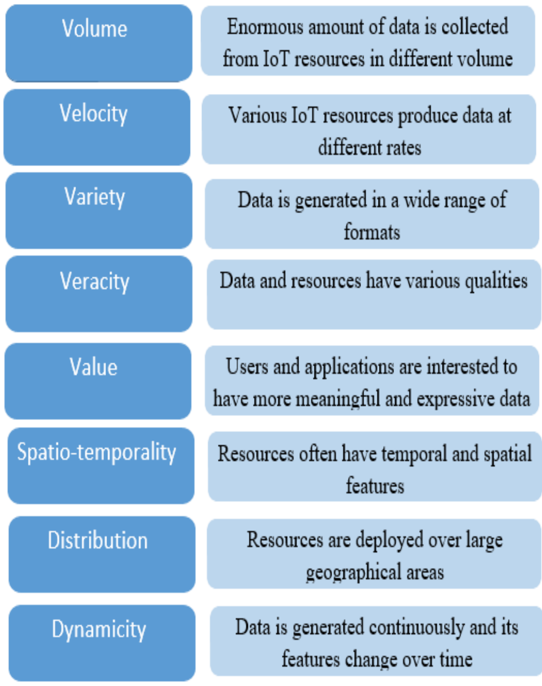

Fig. 2 Characteristics of IoT Data

IoT data does not only have big data characteristics, but IoT data has also dynamicity, distribution and spatio-temporality characteristics [12]. Different IoT applications have different requirements and collect various types of data. For example, different services communicate directly with mobile devices to track their locations in smart connected vehicles and traffic monitoring applications. In real-time railway application, GPS unit is associated with each train to enable users finding departures and arrivals in a real-time.

5. Data Fusion and Filtration

The term data fusion, although widely used, does not have a definitive meaning, varying according to the context in which it is applied. According to Nakamura et al. (2007), several terms (data fusion, sensor fusion and information fusion) are used to describe some aspect of fusion [13]. Still, the terms data fusion and information fusion are accepted and used interchangeably. It is also important to make the distinction between data fusion, sensor fusion and data aggregation clear. Data fusion and information fusion are interchangeable terms, which deal with the fusion of data from any source, including sensors. Data aggregation focuses on reducing data from any type of source, including sensors.

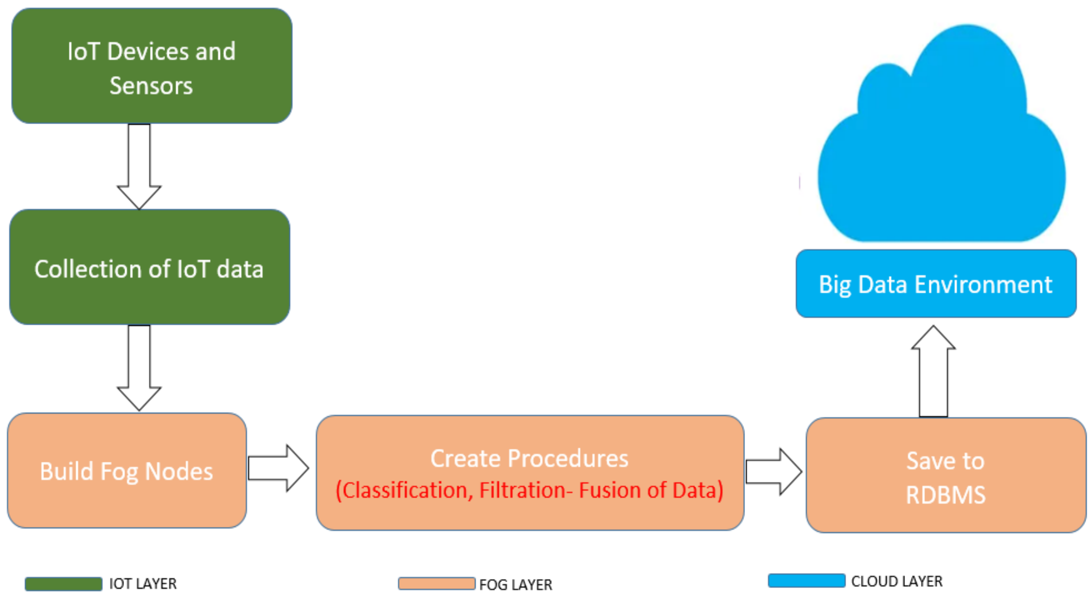

Fig. 3 IoT Data Filtration and Fusion at Fog Layer

Fusion of sensors (or multisensory fusion) is a subgroup that deals with the union of sensory sources. Multisensory integration, on the other hand, applies to the use of sensory data and associated sources (such as databases) to make inferences and interact with the environment. Therefore, sensor fusion is contained at the intersection between multisensory integration and data fusion.

Even with data conversion, it is necessary to ensure that the data is within normal operating parameters. Simple filtering by specifying the minimum and maximum limits of each sensor prevents a typical data (outliers) from causing erroneous decisions at the next level of the architecture. Outliers are data that deviate from the expected value and can be caused by reading errors or errors in the experiment. Therefore, it is important that outliers are properly identified, because otherwise, important information may end up being mistakenly discarded. This filtering step can also be used to identify problematic sensors that need repair or replacement.

For example, in basic filtering, the minimum and maximum limits should be set with values suitable for the region's climate and, if the temperature deviates from the values defined by the user, it will be removed and the sensor can be marked as potentially defective.

Filtering identification and removing outliers is used to compare data from different sensors, that perform a comparison and identify the data that deviates from the standard. Characterization of data fusion methods can also be known as inference and reasoning, Probabilistic methods, Statistical methods, Knowledge-based methods.

Bayesian network, maximum likelihood estimation methods, inference theory, Kalman filtering, and so on comes under the probabilistic methods. The covariance, cross variance, and other statistical analyses comes under Statistical methods, genetic algorithms, Knowledge-based methods include artificial neural networks, and fuzzy logic. Appropriate data fusion methods are being used depending upon the specific problem. The most used method is Kalman filter. Bayesian is also very famous.

6. Multilevel Fusion Architecture

As per definition of data fusion, it aggregates and integrates all sensor data to get accurate and meaningful data while eliminating not required and useless data. So to understand the achievement and benefit of data fusion, it is also important to know about the architecture where it is being implemented. The data fusion multilevel architecture based on the classifications presented by Dasarathy (1997) is very famous and based on the abstraction levels of data input and output. This can be considered as an extension of the classification based on the levels of abstraction of data, but with greater granularity. It expands the concepts of data, characteristic and decision (equivalent to low, medium and high levels) in five categories [14].

- **Data Input - Data Output (DAI-DAO):** The merger of this level deals with pure data like input and output, possibly resulting in more reliable and accurate values. Image and signal processing algorithms may be implemented at this phase.

- **Data Input - Feature Output (DAI-FEO):** The merger of this level uses pure data to apprehend characteristics or attributes of an entity.
- **Feature Input - Feature Output (FEI-FEO):** Merging with input and output features allows you to refine a feature or create a new one.
- **Feature Input - Decision Output (FEI-DEO):** This fusion level uses a group of characteristics to generate a decision.
- **Decision Input - Decision Output (DEI-DEO):** And at this level, decisions are merged to increase the accuracy of a decision or generate a new one.

In distributed system of IoT ecosystem, the fusion of data work is split at different units like at sensors level (Data Collection), Fog Nodes level (Processing Units) and data Centre (Big Storage at Cloud level). The three level IoT data Processing architecture that is hierarchical in nature and applied at various levels of integrated architecture. This can be shown in the figure 3.

7. Case Study: Three-Level Data Fusion for Smart Health Care

The manual process of healthcare can be replaced by smart healthcare automated system. That will minimize the time and energy for better health services. The timely detection of patient condition and its proper action will help in critical services of health sector. This is the reason the use of IoT technologies in this sector expanding every day. IoT is supporting core functions and services of a health service and center. Hospitals are being change in interconnected systems of smart devices. For example, the healthcare body sensors to monitor the patients' critical conditions would generate massively voluminous data. These sensed data must be processed and analyze in real-time, in order to develop knowledge and decision making on time. So the data processing composite different functions, likes de-noising of data, finding of outlier in data, checking missing data and data aggregation.

The data fusion is implemented to handle these various sensor data challenges in different level. Different level of fusion can be shown in figure. 4.

At the lowest level in a smart hospital may have smart beds with multiple sensing devices that collects body position, blood pressure and temperature, heart rhythm, oxygen level, level of fatigue, rate of respiration and many more metrics [15].

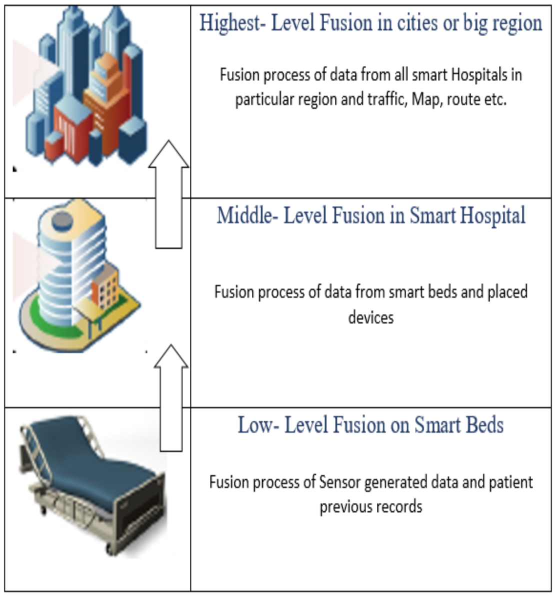

Fig. 4 Data Fusion at Three-level in smart Healthcare

As the main goal at this level, these devices may detect these critical conditions of a particular patient. Based on the collected data and patient previous records, devices can generate and send critical alert to next level likes hospital staff and other responsible person. This way, data fusion will be performed here close to the data sources by integrating reading of devices and previous information.

The middle level implements and uses data fusion at comparatively large level. At this level, it detects potential disease outbreaks by lower level bed sensors data and patient locations. This way, if there is potential disease spread, the middle level turns out to be aware of rooms and specific floors. By knowing the floors and rooms within the hospital, the demand of healthcare will be increased (including epidemic diseases). Middle level organizes the staff effectively and efficiently (e.g. by allocating more health workers to a specific floor, where currently an increased demand is being observed).

For the same, Middle level is depending upon the Lower-level knowledge, individual sensor devices should be mapped for specific rooms and floors where the containment is being reported. Not only efficiently managing hospital personnel by assigning them according to current demand is possible but also system is able to detect a spread of sudden infectious diseases by using data fusion method over these multiple data sources.

The whole managed urban area uses the data fusion which use collected data from the highest level whose work is to collect the data from the hospitals. First of all, the global real time view on the current way of utilizing the hospitals and other healthcare institutions and managing emergency vehicles is possible by using cloud platform. Moreover, highest level can assimilate streaming information from the hospitals with the help of GPS locations when there is an emergency of traffic routes and location of hospital which is located nearby. Resulting, in the emergency the vehicle can have the exact location of the hospital nearby, which has to facility to cater and provide the first aid in time. Second, the prevention of an outbreak of an epidemic disease can be correlated on the regional scale quickly by identifying and analyzing the information which is coming from the different hospitals.

8. Conclusion

In a nutshell, this paper presents a hybrid approach in that data is gone through filtration and fusion process in Fog before sending to local cloud. We explored techniques, implementation and their benefits for local data processing of IoT devices. After the collection of data, we focused on the analysis of outlier detection techniques and resulted in the choice of the most appropriate method for the collected data, aimed at the domain of different sectors of IoT infrastructure. A multilevel data fusion architecture provides an improvement in the accuracy of the collected data, allowing the detection of events and decision making. This can be implemented and applied in many domains of smart environments likes smart cities, smart health, smart agriculture etc. The data fusion structure is decoupled from the infrastructure, allowing its application to be concentrated in the sensor nodes themselves or entirely in the cloud, with the nodes sending only the raw data. This approach is feasible and reduces computation and storage. Also it minimizes delay in network transmission. A more complex and thorough experimental validation and evaluation of this discussed approach will be done to get publishable results in the future.

Acknowledgments

The authors would like to thank the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University, Alkharj, Saudi Arabia for the assistance.

References

[1] https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/(Access Available on 23 Nov,2020)

[2] Data volume of IoT connected devices worldwide 2019 and 2025,Published by S. O'Dea, Oct 26, 2020, https://www.statista.com/statistics/1017863/worldwide-iot-connected-devices-data-size/ (Access Available on 23 Nov,2020)

[3] Ahmad S., Afzal M.M. (2020) Deployment of Fog and Edge Computing in IoT for Cyber-Physical Infrastructures in the 5G Era. In: Karrupusamy P., Chen J., Shi Y. (eds) Sustainable Communication Networks and Application. ICSCN 2019. Lecture Notes on Data Engineering and Communications Technologies, vol 39. Springer, Cham

[4] Callegaro, R.; Montez, C .; Pinto, AR; Moraes, R.(2014) An Architecture for Data Fusion and Outlier Detection in Low Cost Sensors of Wireless Sensor Networks. Proceedings of the II Communication Workshop on Critical Embedded Systems - WoCCES , p. 3–16, 2014.

[5] N. Wichit, "Multisensor data fusion model for activity detection," 2014 Twelfth International Conference on ICT and Knowledge Engineering, Bangkok, 2014, pp. 54-59, doi: 10.1109/ICTKE.2014.7001535.

[6] Jun Wang, Yuan Gao, Chenjian Ran and Yinlong Huo, "State estimation with two-level fusion structure," 2015 International Conference on Estimation, Detection and Information Fusion (ICEDIF), Harbin, 2015, pp. 105-109, doi: 10.1109/ICEDIF.2015.7280171.

[7] A. De Paola, P. Ferraro, S. Gaglio, G. L. Re and S. K. Das, "An Adaptive Bayesian System for Context-Aware Data Fusion in Smart Environments," in IEEE Transactions on Mobile Computing, vol. 16, no. 6, pp. 1502-1515, 1 June 2017, doi: 10.1109/TMC.2016.2599158.

[8] Sheldon Bish, Matthew Rohrer, Peter Scheffel, Kelly Bennett, "Multi-sensor fusion development," Proc. SPIE 9831, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR VII, 98310T (12 May 2016); https://doi.org/10.1117/12.2224138

[9] Segall, R. S., & Niu, G. (2018). Big Data and Its Visualization With Fog Computing. International Journal of Fog Computing (IJFC), 1(2), 51-82.

[10] Sultan Ahmad, Mohammad Mazhar Afzal and Abdullah Alharbi (2020). Big Data Analtics with Fog Computing in integrated Cloud Fog and IoT Architecture for Smart Devices, Internationl Journal of Computer Science and Network Security, Vol. 20 No. 6 pp. 171-177;

[11] Yuri Demchenko, Paola Grosso, Cees De Laat, and Peter Membrey. Addressing big data issues in scientic data infrastructure. In Collaboration Technologies and Systems (CTS), 2013 International Conference on, pages 48{55. IEEE, 2013.

[12] Yasmin Fathy, Payam Barnaghi, and Rahim Tafazolli. Large-Scale Indexing, Discovery, and Ranking for the Internet of Things (IoT). ACM Computin Surveys (CSUR),51(2):29, 2018.
[13] Eduardo F. Nakamura, Antonio A. F. Loureiro, and Alejandro C. Frery. 2007. Information fusion for wireless sensor networks: Methods, models, and classifications. ACM Comput. Surv. 39, 3 (2007), 9–es. DOI:https://doi.org/10.1145/1267070.1267073
[14] B. V. Dasarathy, "Sensor fusion potential exploitation-innovative architectures and illustrative applications," in Proceedings of the IEEE, vol. 85, no. 1, pp. 24-38, Jan. 1997, doi: 10.1109/5.554206
[15] Alexandros Pantelopoulos and Nikolaos G Bourbakis. 2010. A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 40, 1 (2010), 1–12.