Efficient Information Fusion from Environmental Sensors for Near Real-Time IoT Analytics

Maha Ibrahim

Department of International Business Management, Tashkent State University of Economics, Uzbekistan

Email: M.abdelazim@tsue.uz

Abstract

This research paper explores the world of nearly real time analytics by focusing on methods of combining information obtained from Environmental Sensor data. The study utilized a customized setup consisting of three arrays of sensors connected to Raspberry Pi devices. It. Analyzed a dataset that encompassed various environmental conditions. By utilizing the Random Forest algorithm this research investigated how sensor readings, including temperature, humidity, LPG concentrations, smoke, light intensity and motion detection can be fused together. The methodology used cross-validation to ensure model training while visually presenting the intricate relationships, between environmental parameters. The results demonstrated the performance of the Random Forest model through visualizations showing Out of Bag (OOB) error rates and a comparative analysis of machine learning classifiers. The findings shed light on the potential of combining information from sensors to enable reliable predictions using Environmental Sensor data. This provides a foundation for advancements in analytics and applications related to environmental monitoring.

Keywords: Environmental sensor; Information fusion; Internet of Things (IoT); Real-time data processing Sensor data integration; Fusion algorithms; real-time analytics; Multi-sensor information fusion

1. Introduction

The integration of environmental sensors within the framework of the Internet of Things (IoT) has propelled a paradigm shift in data acquisition and analytics. This fusion has enabled the collection of vast and diverse sets of data, nurturing the promise of near real-time insights into our surroundings [1]. The synergy between these technologies has not only revolutionized data acquisition but has also ushered in a new era of analytics, allowing for a deeper understanding of our environment and its dynamics [2]. However, the efficient extraction of meaningful and actionable insights from this deluge of data remains a significant challenge. The sheer volume, heterogeneity, and often disparate nature of data generated by environmental sensors demand sophisticated information fusion techniques. This paper is dedicated to exploring the pivotal role of efficient information fusion methodologies in harnessing the potential of environmental sensor data for near real-time IoT analytics [3-6].

By delving into the complexities of data fusion, this study aims to elucidate the critical mechanisms and strategies employed in amalgamating data from diverse environmental sensors. Furthermore, it seeks to elucidate how this fusion optimizes the analytics process, enabling the extraction of timely, accurate, and comprehensive insights [7]. Throughout this discourse, we will navigate through the landscape of environmental sensor networks, delve into the intricacies of data fusion methodologies, and explore the implications and applications of near real-time analytics in diverse domains [8-11]. By illuminating the significance of efficient information fusion techniques, this paper endeavors to contribute to the evolving field of IoT analytics and advance our capabilities in understanding and responding to environmental phenomena in a more timely and insightful manner [12-15].

Doi: https://doi.org/10.54216/NIF.020201
Received: May 21, 2023 Accepted: October 03, 2023

2. Related Works

This section serves as a critical cornerstone in contextualizing the current landscape of research and advancements in the field of efficient information fusion from environmental sensors for near real-time IoT analytics.

Himeur et al. [15] presented a comprehensive overview of data fusion strategies targeted at enhancing energy efficiency in buildings. Their work highlighted challenges and novel directions in this domain, providing valuable insights into fusion methodologies aimed at optimizing energy usage within built environments. Erhan et al. [16] conducted a thorough review focusing on smart anomaly detection in sensor systems. This multifaceted analysis offered diverse perspectives, elucidating approaches to detect anomalies in sensor data, thereby contributing to the enhancement of data reliability and system robustness. In addition, Gravina et al. [17] delved into the complexities of multi-sensor fusion within body sensor networks, outlining the state-of-the-art techniques and underscoring research challenges. Their work contributed to the understanding of amalgamating data from diverse sensors for healthcare applications. Ouhami et al. [18] conducted a survey on the integration of computer vision, IoT, and data fusion for crop disease detection using machine learning. Their research explored ongoing advancements in remote sensing techniques, particularly in agriculture, paving the way for more accurate and timely disease detection in crops.

Moreover, Munir et al. [19] focused on the integration of artificial intelligence and data fusion at the edge, highlighting their significance in aerospace and electronic systems. Their insights shed light on the potential of edge computing in enabling efficient real-time data fusion for various applications. Muzammal et al. [20] proposed a multi-sensor data fusion approach for medical data obtained from body sensor networks. Their ensemble-based methodology showcased advancements in medical data analysis, contributing to improved diagnostics and healthcare services. Wang et al. [21] explored environmental monitoring utilizing the fog computing paradigm and IoT, illustrating its potential in gathering and processing environmental data. Their work highlighted the role of fog computing in facilitating real-time analytics for environmental monitoring systems. Din et al. [22] introduced a cluster-based data fusion technique for analyzing big data in wireless multi-sensor systems. Their research presented a novel approach to handle large volumes of data efficiently, enhancing the reliability of wireless sensor networks.

Lin et al. [23] investigated multi-sensor fusion in medical human-robot interaction scenarios within body sensor networks. Their research focused on enhancing the interaction between humans and robots in medical settings through advanced sensor fusion techniques. Tsanousa et al. [24] conducted a review encompassing multisensor data fusion solutions in smart manufacturing, elucidating evolving systems and trends. Their work highlighted the role of data fusion in optimizing manufacturing processes within smart environments. Kumar et al. [25] provided a comprehensive survey on machine learning algorithms tailored for wireless sensor networks. Their study outlined various machine learning techniques, showcasing their potential applications and advancements in wireless sensor networks. Iurcev et al. [26] proposed improved automated methods for near real-time mapping applications in the environmental domain. Their research focused on enhancing mapping techniques, particularly in environmental monitoring, contributing to more efficient and accurate real-time mapping methodologies.

3. Methodology

In this section 3, we focus is on delineating the systematic approach employed in this study to investigate and analyze the efficient information fusion from environmental sensors for near real-time IoT analytics.

The Random Forest algorithm, a prominent ensemble learning technique in machine learning, operates on the principle of constructing multiple decision trees and amalgamating their outputs to make robust and accurate predictions. This method involves creating a multitude of decision trees, each trained on a distinct subset of the dataset and utilizing a random selection of features (refer to Algorithm 1). Through this process, the algorithm mitigates overfitting tendencies often encountered in single decision tree models. In the context of our study involving Environmental Sensors, the Random Forest algorithm serves as a pivotal tool in information fusion and learning from sensor data. By employing this algorithm, we aim to harness the diverse readings obtained from these sensors, such as temperature, humidity, CO levels, LPG concentrations, smoke detection, light intensity, and motion detection. The algorithm is adept at accommodating varied types of data and effectively fusing information from these disparate sources. It facilitates the creation of an integrated model that can learn patterns, relationships, and correlations existing among these sensor readings, thereby enabling us to derive comprehensive insights into environmental conditions.

The Random Forest algorithm's strength lies in its ability to handle large volumes of data, mitigate overfitting, and provide robust predictions by aggregating the outputs of multiple decision trees. In our application, it allows us to

Doi: https://doi.org/10.54216/NIF.020201
Received: May 21, 2023 Accepted: October 03, 2023

effectively fuse information obtained from Environmental Sensors, enhancing the accuracy and reliability of our analytics by learning from the combined information captured by these sensors in near real-time. This approach aids in identifying intricate patterns and relationships within the environmental data, ultimately contributing to a more comprehensive understanding of the monitored surroundings.

Algorithm 1: Random Forest Pseudocode

```
To generate c classifiers:
for i = 1 to c do
    Arbitrarily sample the training data D with substitution to produce D_i
    Make a root node, N_i containing D_i

    Call BuildTree (N_i)
end for


BuildTree(N):
  if N comprises examples of only one class then
     return
  else
     Arbitrarily choose x% of the likely splitting features in N
     Choose the feature F with the maximum information gain to split on
     Make f child nodes of N, N_1, ..., N_f, where F has f likely values (F_1, ..., F_f)
    for i = 1 to f do
         Set the contents of N_i to D_i, where D_i is all examples in N that match F_i

         Call BuildTree (N_i)
     end for
  end if.
```

To ensure the robustness and reliability of our model trained on the Environmental Sensor data, we implement a 5-fold cross-validation technique. This approach involves partitioning the dataset into five equal subsets while sequentially designating each subset as the validation set and the remaining four as the training set. Through this iterative process, the model undergoes training and validation five times, each time utilizing a different partition for validation while training on the other four partitions. By systematically rotating through these subsets, we effectively mitigate biases and variance in our model's performance evaluation. This method allows us to assess the model's generalization capabilities across various subsets of the dataset, enhancing its adaptability to diverse environmental conditions and ensuring that the derived insights and predictions are reliable and consistent across different segments of the data.

4. Experimental Design

This section elucidates the specific design, setup, and execution of experiments conducted to validate the efficacy and performance of the developed information fusion framework for near real-time IoT analytics.

In this study, we present a case study that encompasses the acquisition of data from a trio of identical, specifically crafted sensor arrays constructed on breadboards. Each array was intricately linked to a Raspberry Pi device and stationed in distinct physical locations characterized by diverse environmental conditions. The data collection process involved these three individuals IoT devices, each consistently gathering seven distinct readings from four sensors at regular intervals. These sensor readings encapsulated crucial environmental parameters such as temperature, humidity, carbon monoxide (CO), liquid petroleum gas (LPG), smoke, light intensity, and motion detection. Spanning from 07/12/2020 00:00:00 UTC to 07/19/2020 23:59:59 UTC, the dataset encompasses a substantial volume, comprising 405,184 rows of meticulously gathered data. To ensure seamless data transmission, the sensor readings, coupled with unique device identifiers and timestamps, were encapsulated as a unified message adhering to the ISO standard of

Doi: https://doi.org/10.54216/NIF.020201
Received: May 21, 2023 Accepted: October 03, 2023

Message Queuing Telemetry Transport (MQTT) network protocol. In Table 1, we provide summary statistics for our case study.

Table 1: Summary of descriptive statistics for environmental sensory data.

	co	humidity	lpg	smoke	temp
count	405184	405184	405184	405184	405184
mean	0.004639	60.51169	0.007237	0.019264	22.45399
std	0.00125	11.36649	0.001444	0.004086	2.698347
min	0.001171	1.1	0.002693	0.006692	0
25%	0.003919	51	0.006456	0.017024	19.9
50%	0.004812	54.9	0.007489	0.01995	22.2
75%	0.005409	74.3	0.00815	0.021838	23.6
max	0.01442	99.9	0.016567	0.04659	30.6

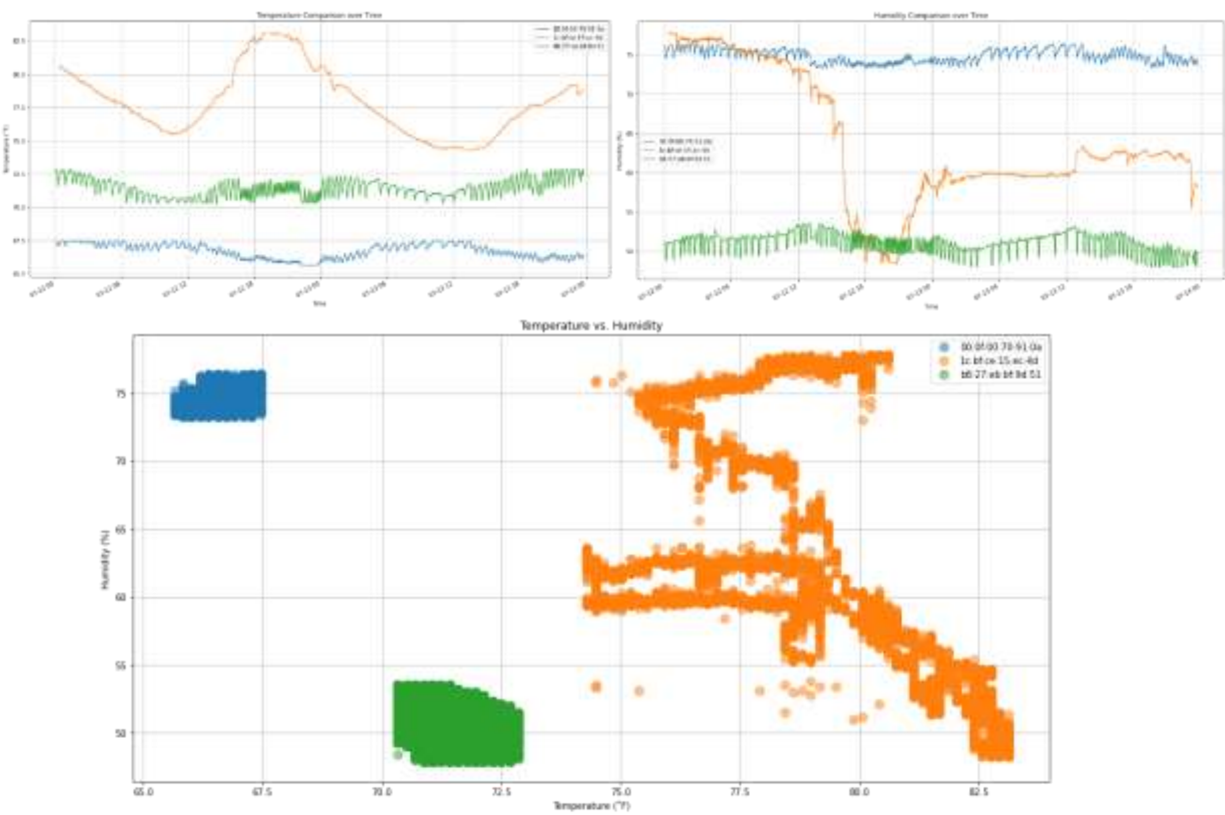

Figure 1: Cyclic Variation of Temperature and Humidity Over Time: A Visual Representation of Environmental Sensor Data

5. Results and Discussion

This section encapsulates the empirical findings, analyses, and their interpretations derived from the conducted experiments, shedding light on the performance and implications of the developed information fusion framework for near real-time IoT analytics.

In Figure 1 of our results, we present a comprehensive visualization depicting the temporal flow and correlation between temperature and humidity readings obtained from the Environmental Sensor data. This visualization

Doi: https://doi.org/10.54216/NIF.020201
Received: May 21, 2023 Accepted: October 03, 2023

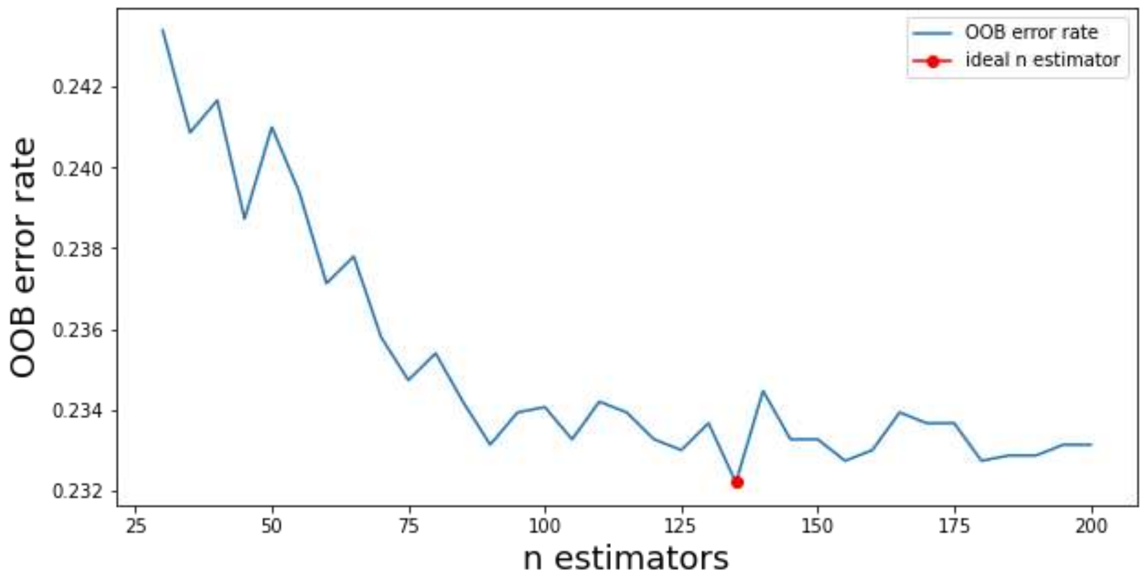

Figure 2: Out-of-Bag (OOB) Error Rate Evolution with Increasing Decision Trees in the Random Forest Ensemble

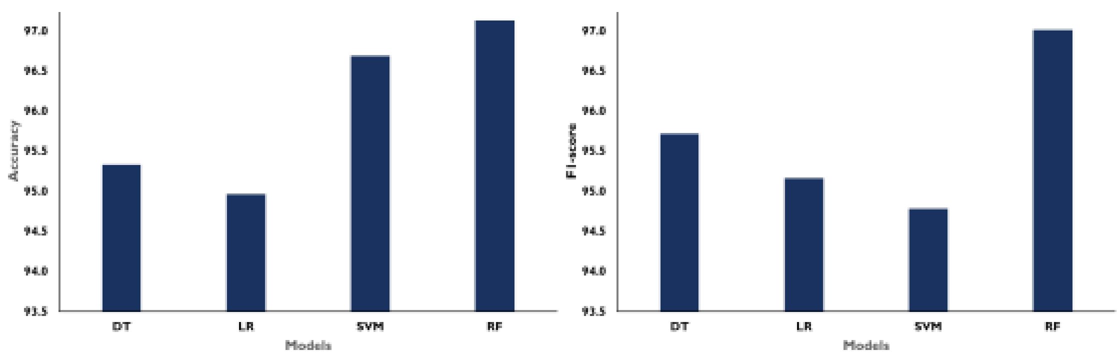

Figure 3: Performance Evaluation of Machine Learning Classifiers on Environmental Sensor Data: Accuracy,, and F1-score Metrics Comparison

encapsulates the dynamic interplay between these two crucial environmental parameters over the specified timeframe. By plotting these variables against time, we aim to elucidate their trends, patterns, and potential correlations, providing a clear and intuitive understanding of their relationship. This visualization facilitates the identification of potential dependencies or associations between temperature and humidity levels, offering insights into how changes in one parameter may impact the other. Through this graphical representation, we enhance the interpretability of the data, enabling stakeholders to grasp the concurrent fluctuations and potential interactions between temperature and humidity within the monitored environment.

``in Figure 2, we visualize the Out-of-Bag (OOB) error rate visualization, a critical metric derived from the Random Forest model trained on our Environmental Sensor dataset. This graphical representation elucidates the model's performance during training by plotting the OOB error rate against the number of decision trees utilized in the ensemble. The OOB error rate serves as a crucial indicator of the model's predictive accuracy, reflecting how effectively the model generalizes to unseen data. This visualization allows for a comprehensive assessment of the model's learning curve, providing insights into the convergence of the error rate as more decision trees are incorporated into the ensemble. By observing this graphical representation, we gain valuable insights into the model's stability, overfitting tendencies, and its ability to make accurate predictions as it learns from the Environmental Sensor data.

In Figure 3, we present a comparative analysis showcasing the performance of multiple machine learning classifiers applied to our Environmental Sensor dataset. This visualization encapsulates the accuracy, and F1-score metrics

Doi: https://doi.org/10.54216/NIF.020201
Received: May 21, 2023 Accepted: October 03, 2023

attained by various classifiers, namely Random Forest, Support Vector Machines (SVM), Logistic regression, and Decision Tree. By juxtaposing the performance metrics of these classifiers, we aim to provide a comprehensive overview and comparative assessment of their predictive capabilities in handling the intricacies of the sensor data. This visualization enables stakeholders to discern the strengths and weaknesses of each classifier, facilitating an informed selection of the most suitable model for deriving insights and making predictions from the Environmental Sensor dataset.

6. Conclusion

This study embarked on a comprehensive exploration of efficient information fusion methodologies from Environmental Sensor data for near real-time IoT analytics. Through the amalgamation of diverse sensor readings encompassing temperature, humidity, CO levels, LPG concentrations, smoke, light intensity, and motion detection, our approach leveraged the Random Forest algorithm to fuse and glean insights from these disparate data sources. The application of 5-fold cross-validation ensured robustness in model training, while visualizations depicted the dynamic interplay between environmental parameters, offering a nuanced understanding of their relationships. Furthermore, our results demonstrated the Random Forest model's performance through visualization of the Out-of-Bag (OOB) error rate and a comparative analysis of multiple machine learning classifiers. The insights derived from this study not only showcase the potential of the Random Forest algorithm in facilitating comprehensive data fusion but also emphasize the importance of employing robust methodologies for accurate and reliable predictions.

References

[1] Alam, Furqan, Rashid Mehmood, Iyad Katib, Nasser N Albogami, and Aiiad Albeshri. 2017. "Data Fusion and IoT for Smart Ubiquitous Environments: A Survey." Ieee Access 5: 9533–54.

[2] Ding, Wenxiu, Xuyang Jing, Zheng Yan, and Laurence T Yang. 2019. "A Survey on Data Fusion in Internet of Things: Towards Secure and Privacy-Preserving Fusion." Information Fusion 51: 129–44.

[3] Kumar, K Kranthi, E Ramaraj, and P Geetha. 2023. "Multi-Sensor Data Fusion for an Efficient Object Tracking in Internet of Things (IoT)." Applied Nanoscience 13 (2): 1355–65.

[4] Krishnamurthi, Rajalakshmi, Adarsh Kumar, Dhanalekshmi Gopinathan, Anand Nayyar, and Basit Qureshi. 2020. "An Overview of IoT Sensor Data Processing, Fusion, and Analysis Techniques." Sensors 20 (21): 6076.

[5] Munir, Arslan, Jisu Kwon, Jong Hun Lee, Joonho Kong, Erik Blasch, Alexander J Aved, and Khan Muhammad. 2021. "FogSurv: A Fog-Assisted Architecture for Urban Surveillance Using Artificial Intelligence and Data Fusion." IEEE Access 9: 111938–59.

[6] Qiu, Sen, Hongkai Zhao, Nan Jiang, Zhelong Wang, Long Liu, Yi An, Hongyu Zhao, Xin Miao, Ruichen Liu, and Giancarlo Fortino. 2022. "Multi-Sensor Information Fusion Based on Machine Learning for Real Applications in Human Activity Recognition: State-of-the-Art and Research Challenges." Information Fusion 80: 241–65.

[7] Palumbo, Filippo, Claudio Gallicchio, Rita Pucci, and Alessio Micheli. 2016. "Human Activity Recognition Using Multisensor Data Fusion Based on Reservoir Computing." Journal of Ambient Intelligence and Smart Environments 8 (2): 87–107.

[8] Shamshiri, Redmond R, Iva Bojic, Eldert van Henten, Siva K Balasundram, Volker Dworak, Muhammad Sultan, and Cornelia Weltzien. 2020. "Model-Based Evaluation of Greenhouse Microclimate Using IoT-Sensor Data Fusion for Energy Efficient Crop Production." Journal of Cleaner Production 263: 121303.

[9] Pang, Zhibo, Qiang Chen, Weili Han, and Lirong Zheng. 2015. "Value-Centric Design of the Internet-of-Things Solution for Food Supply Chain: Value Creation, Sensor Portfolio and Information Fusion." Information Systems Frontiers 17: 289–319.

[10] Hall, David, Chee-Yee Chong, James Llinas, and Martin Liggins II. 2017. Distributed Data Fusion for Network-Centric Operations. Crc Press.

[11] Ounoughi, Chahinez, and Sadok Ben Yahia. 2023. "Data Fusion for ITS: A Systematic Literature Review." Information Fusion 89: 267–91.

[12] Li, Xianzhi, Qiao Yu, Bander Alzahrani, Ahmed Barnawi, Ahmed Alhindi, Daniyal Alghazzawi, and Yiming Miao. 2021. "Data Fusion for Intelligent Crowd Monitoring and Management Systems: A Survey." IEEE Access 9: 47069–83.

[13] Cheng, Tao, Giovanni C Migliaccio, Jochen Teizer, and Umberto C Gatti. 2013. "Data Fusion of Real-Time Location Sensing and Physiological Status Monitoring for Ergonomics Analysis of Construction Workers." Journal of Computing in Civil Engineering 27 (3): 320–35.

Doi: https://doi.org/10.54216/NIF.020201
Received: May 21, 2023 Accepted: October 03, 2023

[14] Kenda, Klemen, Blaž Kažič, Erik Novak, and Dunja Mladenić. 2019. "Streaming Data Fusion for the Internet of Things." Sensors 19 (8): 1955.
[15] Himeur, Yassine, Abdullah Alsalemi, Ayman Al-Kababji, Faycal Bensaali, and Abbes Amira. 2020. "Data Fusion Strategies for Energy Efficiency in Buildings: Overview, Challenges and Novel Orientations." Information Fusion 64: 99–120.
[16] Erhan, Laura, M Ndubuaku, Mario Di Mauro, Wei Song, Min Chen, Giancarlo Fortino, Ovidiu Bagdasar, and Antonio Liotta. 2021. "Smart Anomaly Detection in Sensor Systems: A Multi-Perspective Review." Information Fusion 67: 64–79.
[17] Gravina, Raffaele, Parastoo Alinia, Hassan Ghasemzadeh, and Giancarlo Fortino. 2017. "Multi-Sensor Fusion in Body Sensor Networks: State-of-the-Art and Research Challenges." Information Fusion 35: 68–80.
[18] Ouhami, Maryam, Adel Hafiane, Youssef Es-Saady, Mohamed El Hajji, and Raphael Canals. 2021. "Computer Vision, IoT and Data Fusion for Crop Disease Detection Using Machine Learning: A Survey and Ongoing Research." Remote Sensing 13 (13): 2486.
[19] Munir, Arslan, Erik Blasch, Jisu Kwon, Joonho Kong, and Alexander Aved. 2021. "Artificial Intelligence and Data Fusion at the Edge." IEEE Aerospace and Electronic Systems Magazine 36 (7): 62–78.
[20] Muzammal, Muhammad, Romana Talat, Ali Hassan Sodhro, and Sandeep Pirbhulal. 2020. "A Multi-Sensor Data Fusion Enabled Ensemble Approach for Medical Data from Body Sensor Networks." Information Fusion 53: 155–64.
[21] Wang, Wendong, Cheng Feng, Bo Zhang, and Hui Gao. 2019. "Environmental Monitoring Based on Fog Computing Paradigm and Internet of Things." IEEE Access 7: 127154–65.
[22] Din, Sadia, Awais Ahmad, Anand Paul, Muhammad Mazhar Ullah Rathore, and Gwanggil Jeon. 2017. "A Cluster-Based Data Fusion Technique to Analyze Big Data in Wireless Multi-Sensor System." IEEE Access 5: 5069–83.
[23] Lin, Kai, Yihui Li, Jinchuan Sun, Dongsheng Zhou, and Qiang Zhang. 2020. "Multi-Sensor Fusion for Body Sensor Network in Medical Human--Robot Interaction Scenario." Information Fusion 57: 15–26.
[24] Tsanousa, Athina, Evangelos Bektsis, Constantine Kyriakopoulos, Ana Gómez González, Urko Leturiondo, Ilias Gialampoukidis, Anastasios Karakostas, Stefanos Vrochidis, and Ioannis Kompatsiaris. 2022. "A Review of Multisensor Data Fusion Solutions in Smart Manufacturing: Systems and Trends." Sensors 22 (5): 1734.
[25] Kumar, D Praveen, Tarachand Amgoth, and Chandra Sekhara Rao Annavarapu. 2019. "Machine Learning Algorithms for Wireless Sensor Networks: A Survey." Information Fusion 49: 1–25.
[26] Iurcev, M, F Pettenati, P Diviacco, and others. 2021. "Improved Automated Methods for near Real-Time Mapping-Application in the Environmental Domain." Bulletin of Geophysics and Oceanography 62 (3): 427–54.

Doi: https://doi.org/10.54216/NIF.020201
Received: May 21, 2023 Accepted: October 03, 2023