Optimal Integration of Data Fusion in Solar Power Analytics: Enhancing Efficiency and Accuracy

Darío González-Cruz[1], Franky Jiménez-García[1], Javier Gamboa-Cruzado[1,2], Edward R. Luna Victoria[1], María Lima Bendezú[3], Reem Attasi[4]

[1] César Vallejo University, Trujillo, Peru
[2] National University of San Marcos, Trujillo, Peru
[3] National University Micaela Bastidas of Apurímac, Apurímac, Peru
[4] Higher Colleges of Technology, United Arab Emirates

Emails: darioucv@ucvvirtual.edu.pe; fjimenezga@ucvvirtual.edu.pe; jgamboac@ucv.edu.pe; erubio@ucv.edu.pe; mlima@unamba.edu.pe; ratassi@hct.ac.ae

Abstract

At the forefront of sustainable energy solutions lies renewable energy, particularly solar power. Nevertheless, the optimization of solar power systems necessitates comprehensive analytics, especially for proactive maintenance fault anticipation. This research evaluates data fusion techniques using both linear and non-linear regression models for predicting faults in solar power plants. The study begins with careful data preparation processes to ensure clean and harmonized data sets that include irradiation, temperature, historical fault records, and yield. Linear regression techniques provide insights into straightforward correlations while non-linear models go deep into complex relationships within the data. The results indicate positive outcomes demonstrating the potential of these fusion techniques as far as improving accuracy in fault prediction is concerned. These findings highlight the importance of refining data preparation prior to any fusion process and recommend further exploration into more advanced fusion methodologies. This paper helps advance proactive maintenance strategies for solar power plants thereby making this source of energy more dependable and resilient.

Keywords: Solar Energy Analytics; Information Fusion; Photovoltaic Systems; Energy Harvesting Analysis; Multi-source Data Fusion; Solar Power Optimization; Machine Learning; Performance Enhancement.

1. Introduction

Because of the rising demand for sustainable and renewable energy sources, there has been a significant shift towards harnessing solar power. In this fresh interest, the efficiency and accuracy of solar power analytics have become critical areas for research and development [1-3]. To address these issues, this paper explores optimal data fusion techniques in the field of solar power analytics, with an aim to improve the accuracy and effectiveness of energy evaluation models [4].

Renewable energy is increasingly hinged on solar power which offers a promising means for meeting global growing demands for energy while also mitigating environmental impact [5-7]. However, it is crucially important to optimize various datasets that come from different sources like weather conditions, solar panel performance, geographical positioning, and energy production trends so as to maximize solar potential. Nevertheless, integrating these separate datasets harmoniously is a major challenge that calls for efficient data fusion methodologies [8-10].

This study embarks on an exploration of cutting-edge data fusion techniques and their integration into the realm of solar power analytics. By amalgamating diverse data sources and applying advanced analytical frameworks, the aim is to optimize the precision of solar energy assessments. The overarching objective is to not only enhance the efficiency

Doi: https://doi.org/10.54216/FPA.140217
Received: July 17, 2023 Revised: November 12, 2023 Accepted: January 17, 2024

of energy production but also to elevate the accuracy of predictive models, facilitating informed decision-making in the deployment and management of solar power systems.

In the subsequent sections, this paper will delve into the fundamental principles of data fusion, elucidate its relevance in the context of solar power analytics, discuss various methodologies employed for integration, and critically evaluate their impact on enhancing efficiency and accuracy. Additionally, this study will present case studies and empirical evidence to underscore the efficacy of these techniques in real-world scenarios, further substantiating the significance of optimal data fusion in advancing solar energy analytics.

2. The proposd Methodology

In this section, the methodology employed for enhancing solar power analytics revolves around the strategic integration of advanced data fusion techniques. The crux of this approach lies in amalgamating diverse and multi-level data sources, spanning irradiation, ambient and module temperatures, energy yield, and additional pertinent variables, to create a cohesive and comprehensive understanding of solar energy systems.

Algorithm 1: Solar data preprocessing Algorithm
Input: $D_{\text{train}}, d_i^j = (x_i^j, c_n), d \in D_{\text{train}}, D_{\text{test}}, w_i^j = (w_i^j, c_n), w \in D_{\text{test}}$ and Parameters are Γ and Δ
Output: $\hat{D}, \hat{d} = (x_e^n, c_n), \hat{d} \in \hat{D}$
1: for each instance in $D_{\text{train},}$, identify class do
2: $\tau = \{T_0^{C_m}\}_{m=1}^{n}$;
3: for Each Epoch select subspace in τ, based on size of space z_Δ do
4: $\Delta = \text{Sample}\ (D_{\text{train}}, z_\Delta)\Delta = \{ST_0^{C_m}\}_{m=1}^{n}$
5: end for
6: for Each Epoch select subspace in τ, based on the size of space z_Γ do
7: $\hat{\Gamma} = \text{Sample}\ (D_{\text{train}}, z_{\hat{\Gamma}})\hat{\Gamma} = \{L_0^{C_m}\}_{m=1}^{n}$;
8: end for
9: To make algorithm's robust, we state the ϕ noisy;
10: $\phi = \{\text{Nos}^{C_m}\}_{m=1}^{n}$;
11: $\Gamma = \phi + \hat{\Gamma}$;
Preparing training sensory data;
12: for $i = 1, i \leq n, i++$ do
13: $\Delta = \{ST_0^{C_m}\}_{m=1}^{n}$, based on the z_Δ;
14: for $j = 1, i \leq n, j++$ do
15: $\Gamma = \{L_0^{C_m}\}_{m=1}^{n}$, based on the z_Γ;
16: $\hat{D}_{\text{train}} \leftarrow$ match (Γ, Δ);
Preparing testing sensory data;
17: for $i = 1, i \leq n, i++$ do
18: $\Delta = \{ST_0^{C_m}\}_{m=1}^{n}$, based on the z_Δ;
19: end for
20: for $w = 1, i \leq n, w++$ do
21: $\hat{d} = (w_r^j, c_n), \hat{d} \in \hat{D}_{\text{test}}$, based on the $p = z_\Gamma$;
22: end for
23: $\hat{D}_{\text{test}} \leftarrow$ match $(\hat{d}, \Delta)$;
24: return $\hat{D}_{\text{train}}$ and $\hat{D}_{\text{test}}$;

Prior to engaging in the fusion process aimed at integrating diverse datasets for solar power analytics, a series of meticulous data preparation steps were executed to ensure the cleanliness, consistency, and reliability of the datasets. The preparatory phase constituted a fundamental component of this research, pivotal for fostering the efficacy and accuracy of the fusion process. Initial data collection involved gathering raw data from disparate sources encompassing solar irradiation, ambient and module temperatures, energy yield, geographic positioning, and

Doi: https://doi.org/10.54216/FPA.140217
Received: July 17, 2023 Revised: November 12, 2023 Accepted: January 17, 2024

additional relevant variables. Data acquisition adhered to established protocols and standards to ensure uniformity and reliability across diverse sources [11-13]. Then, after data acquisition, a stringent data cleaning regimen was employed to rectify inconsistencies, errors, and missing values within the datasets. This entailed the identification and removal of outliers, handling of null or erroneous entries through imputation or elimination, and standardization of data formats to maintain coherence and consistency. After that, Alignment and synchronization of temporal and spatial parameters were imperative for harmonizing datasets originating from varying sources and timeframes. This involved timestamp synchronization, geographical alignment, and normalization of temporal intervals to establish a unified timeline across datasets, ensuring their compatibility for fusion. To streamline the fusion process and mitigate redundancy, feature selection techniques were applied to identify pertinent variables contributing significantly to solar power analytics. Feature extraction methodologies were also employed to derive new composite features or reduce dimensionality while retaining essential information. The final phase of data preparation encompassed rigorous quality assurance checks and validation procedures [14-16].

The fusion process involved the application of linear regression techniques to amalgamate diverse datasets encompassing parameters such as irradiation, ambient and module temperatures, historical fault data, energy yield, and other relevant variables. Linear regression, known for its simplicity and interpretability, served as an effective method for integrating these variables and establishing relationships between them.

$$P(t) = a + b \cdot E(t) \tag{1}$$

Following the fusion of cleaned datasets through linear regression, the derived model was employed for predictive analytics pertaining to faults in solar plants. Leveraging the fused dataset, the linear regression model was trained on historical fault data to discern patterns, correlations, and causal relationships between various input variables and the occurrence of faults. Subsequently, this trained model was utilized to predict and anticipate potential fault occurrences within solar power systems.

The data fusion process involved the application of non-linear regression techniques to integrate the cleaned datasets comprising various parameters such as irradiation, ambient and module temperatures, historical fault data, energy yield, and other pertinent variables [17-19]. Non-linear regression methods, known for capturing complex relationships and interactions among variables, were employed to accommodate non-linear patterns and intricate dependencies present within the dataset. This allowed for a more nuanced fusion of data, capturing intricate relationships that might not be adequately modeled by linear techniques.

$$P(t) = aE(t)\left(1 - b\left(T(t) + \frac{E(t)}{800}(c - 20) - 25\right) - d\ln\left(E(t)\right)\right) \tag{2}$$

Subsequent to the fusion of cleaned datasets using non-linear regression, the derived model was utilized for predictive analysis concerning faults in solar plants. Upon training the non-linear regression model on historical fault data, the aim was to uncover intricate patterns, non-linear correlations, and interdependencies among various input variables and the occurrence of faults. This trained non-linear model was then used to predict and anticipate potential fault occurrences within solar power systems, accounting for the nuanced relationships captured during the fusion process [20-25].

3. Results and Discussion

This section presents the empirical outcomes derived from the application of data fusion techniques, specifically linear and non-linear regression methodologies, in the context of predicting faults within solar power plants. The graphical representation in Figure 1 delineates critical parameters associated with solar power analytics, presenting a

Doi: https://doi.org/10.54216/FPA.140217
Received: July 17, 2023 Revised: November 12, 2023 Accepted: January 17, 2024

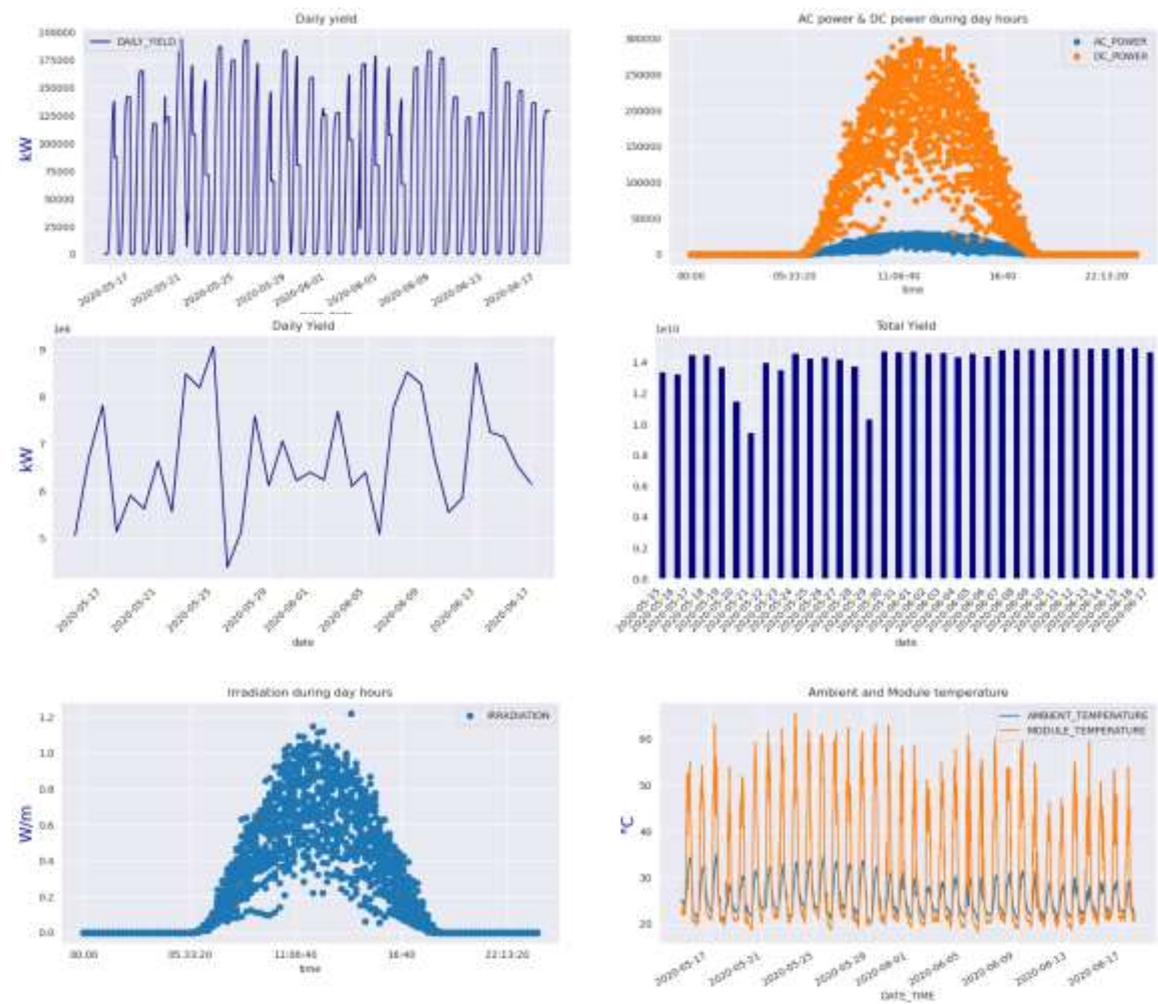

Figure 1: Visual Representation of Solar Power Analytics

comprehensive visualization of Daily Yield versus AC-DC power, Daily and Total Yield, and variations in Irradiation, Ambient, and Module temperatures across different sections – upper, middle, and bottom.

The upper section of Figure 1 illustrates the correlation between Daily Yield and AC-DC power output. This comparative analysis provides insights into the efficiency of power conversion from alternating current (AC) to direct current (DC) concerning the energy yield generated on a daily basis. In the middle segment of Figure 1, the visual representation delineates the trend lines for Daily and Total Yield. This depiction offers a holistic view of the cumulative energy production over time, alongside the daily variations in yield, aiding in trend analysis and long-term performance assessment. The lower part of Figure 1 showcases the variations in Irradiation, Ambient, and Module temperatures across the upper, middle, and bottom sections. These parameters play a crucial role in solar energy generation, influencing the efficiency and performance of photovoltaic systems.

The graphical representation in Figure 2 delineates distinct components of solar power analytics, portraying trend analysis, seasonality, and residual variations across three segments – upper, middle, and bottom. In the upper section of Figure 2, the visualization depicts the identified trend within the solar power data. This segment showcases the long-term pattern or directional movement, aiding in understanding the overarching trajectory of energy production, whether it's an ascending, descending, or stable trend. The middle segment of Figure 2 illustrates the seasonal variations inherent in solar power data. This visual representation captures cyclic patterns or fluctuations occurring at regular intervals, highlighting recurring trends across different time periods, such as daily, monthly, or yearly seasonal effects. The lower part of Figure 2 presents residual variations or irregularities within the solar power data. This

Doi: https://doi.org/10.54216/FPA.140217
Received: July 17, 2023 Revised: November 12, 2023 Accepted: January 17, 2024

segment showcases the deviations or fluctuations that remain after extracting the trend and seasonal components, offering insights into the random or unexpected variations impacting energy production.

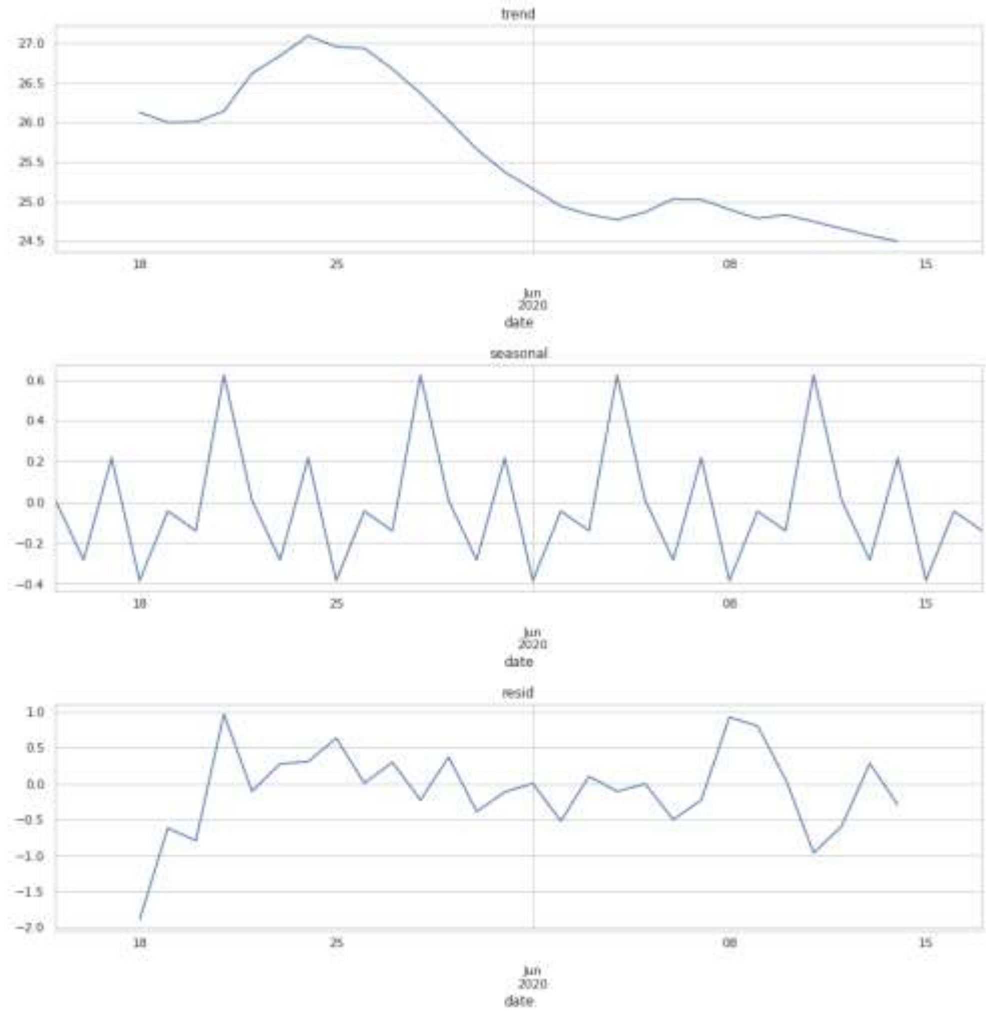

Figure 2: Visual Representation of Analytical Trends in Solar Power Data

Figure 3 presents a correlation map encapsulating the interrelationships and associations among various parameters within solar power analytics. This visual depiction employs a color-coded matrix to illustrate the strength and direction of correlations between different variables, such as irradiation, ambient temperature, module efficiency, and energy yield. The correlation map serves as a comprehensive visual tool, allowing for a quick and intuitive assessment of the degree of linear relationships between key factors impacting solar energy generation. The varying intensities of color denote the magnitude and nature of correlations, aiding in identifying significant correlations crucial for optimizing solar power systems and guiding decision-making processes.

In Figure 4, we present the visualization depicting the predictive outcomes generated by both linear and non-linear regression models. This graphical representation showcases the comparative analysis of predictions derived from the two distinct methodologies. The visualization provides a comprehensive overview of the forecasted fault occurrences within solar power plants, delineating the predictive accuracies and trends yielded by the linear and non-linear models. This visualization serves as a pivotal reference point, enabling a comparative assessment of the predictive performances and guiding further discussions regarding the efficacy and suitability of linear and non-linear regression techniques in fault prediction within solar power plants.

Doi: https://doi.org/10.54216/FPA.140217
Received: July 17, 2023 Revised: November 12, 2023 Accepted: January 17, 2024

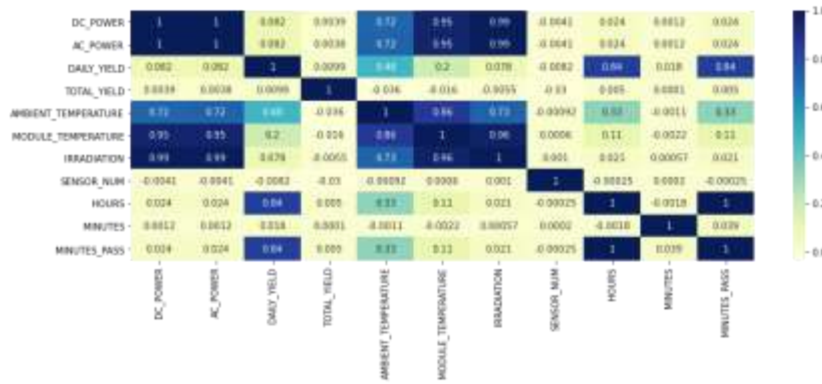

Figure 3: Correlation Map of Solar Power Analytics.

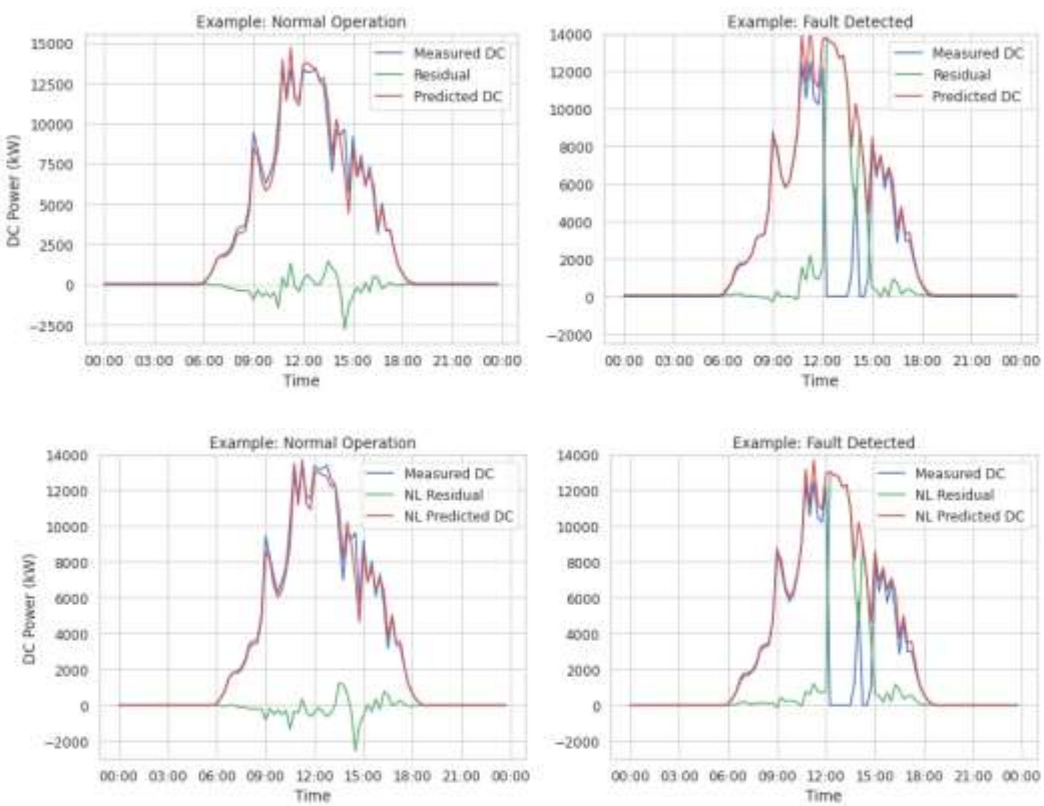

Figure 4: Visualization of Predictions from Linear and Non-linear Models

4. Conclusions

This research has ventured into the realm of data fusion techniques, employing both linear and non-linear regression methodologies to predict faults within solar power plants. The application of these fusion techniques showcased promising results, revealing their potential to enhance the accuracy and predictive capabilities concerning fault anticipation in solar energy systems. The outcomes underscore the significance of meticulous data preparation, emphasizing the critical role it plays in refining and harmonizing diverse datasets for subsequent fusion processes. The comparative analysis between linear and non-linear regression models unveiled the advantages of non-linear techniques in capturing intricate relationships, while linear models provided valuable insights into more straightforward correlations within the data. These findings signify a pivotal step forward in bolstering proactive

Doi: https://doi.org/10.54216/FPA.140217
Received: July 17, 2023 Revised: November 12, 2023 Accepted: January 17, 2024

maintenance strategies within solar power plants, enabling more efficient fault detection and facilitating timely interventions. Nonetheless, this study also acknowledges certain limitations, including the need for further exploration into more sophisticated fusion methodologies and the requirement for comprehensive datasets encompassing a broader spectrum of variables for a more holistic predictive model. Moving forward, the insights garnered from this research pave the way for continued advancements in predictive analytics, fostering a more resilient and reliable solar energy landscape.

References

[1] Nweke, Henry Friday, Ying Wah Teh, Ghulam Mujtaba, and Mohammed Ali Al-garadi. 2019. "Data Fusion and Multiple Classifier Systems for Human Activity Detection and Health Monitoring: Review and Open Research Directions." Information Fusion. https://doi.org/10.1016/j.inffus.2018.06.002.
[2] Muhammad, Ghulam, and M. Shamim Hossain. 2021. "COVID-19 and Non-COVID-19 Classification Using Multi-Layers Fusion From Lung Ultrasound Images." Information Fusion. https://doi.org/10.1016/j.inffus.2021.02.013.
[3] 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.
[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] 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.
[7] Maimaitijiang, Maitiniyazi, Abduwasit Ghulam, Paheding Sidike, Sean Hartling, Matthew Maimaitiyiming, Kyle Peterson, Ethan Shavers, et al. 2017. "Unmanned Aerial System (UAS)-Based Phenotyping of Soybean Using Multi-Sensor Data Fusion and Extreme Learning Machine." ISPRS Journal of Photogrammetry and Remote Sensing 134: 43–58.
[8] Sharma, Pavika, and Pawan Whig. 2023. "Maximizing Solar Energy Utilization through Predictive Machine Learning Techniques." International Journal of Sustainable Development Through AI, ML and IoT 2 (1): 1–13.
[9] Hall, David Lee, and Sonya A H McMullen. 2004. Mathematical Techniques in Multisensor Data Fusion. Artech House.
[10] Himeur, Yassine, Abdullah Alsalemi, Ayman Al-Kababji, Faycal Bensaali, Abbes Amira, Christos Sardianos, George Dimitrakopoulos, and Iraklis Varlamis. 2021. "A Survey of Recommender Systems for Energy Efficiency in Buildings: Principles, Challenges and Prospects." Information Fusion 72: 1–21.
[11] Dong, Jiang, Dafang Zhuang, Yaohuan Huang, and Jingying Fu. 2009. "Advances in Multi-Sensor Data Fusion: Algorithms and Applications." Sensors 9 (10): 7771–84.
[12] Diez-Olivan, Alberto, Javier Del Ser, Diego Galar, and Basilio Sierra. 2019. "Data Fusion and Machine Learning for Industrial Prognosis: Trends and Perspectives towards Industry 4.0." Information Fusion 50: 92–111.
[13] Liu, Zheng, Norbert Meyendorf, and Nezih Mrad. 2018. "The Role of Data Fusion in Predictive Maintenance Using Digital Twin." In AIP Conference Proceedings. Vol. 1949.
[14] Dong, Xin Luna, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Kevin Murphy, Shaohua Sun, and Wei Zhang. 2015. "From Data Fusion to Knowledge Fusion." ArXiv Preprint ArXiv:1503.00302.
[15] James, Alex Pappachen, and Belur V Dasarathy. 2014. "Medical Image Fusion: A Survey of the State of the Art." Information Fusion 19: 4–19.
[16] Ziani, Said. 2023. "Enhancing Fetal Electrocardiogram Classification: A Hybrid Approach Incorporating Multimodal Data Fusion and Advanced Deep Learning Models." Multimedia Tools and Applications, 1–41.
[17] Chen, Bin, Bo Huang, and Bing Xu. 2017. "Multi-Source Remotely Sensed Data Fusion for Improving Land Cover Classification." ISPRS Journal of Photogrammetry and Remote Sensing 124: 27–39.
[18] Ghamisi, Pedram, Behnood Rasti, Naoto Yokoya, Qunming Wang, Bernhard Hofle, Lorenzo Bruzzone, Francesca Bovolo, et al. 2019. "Multisource and Multitemporal Data Fusion in Remote Sensing: A Comprehensive Review of the State of the Art." IEEE Geoscience and Remote Sensing Magazine 7 (1): 6–39.

Doi: https://doi.org/10.54216/FPA.140217
Received: July 17, 2023 Revised: November 12, 2023 Accepted: January 17, 2024

[19] Fawzy, Dina, Sherin M Moussa, and Nagwa L Badr. 2023. “An IoT-Based Resource Utilization Framework Using Data Fusion for Smart Environments.” Internet of Things 21: 100645.
[20] Gutiérrez, Raúl, V\'\ictor Rampérez, Horacio Paggi, Juan A Lara, and Javier Soriano. 2022. “On the Use of Information Fusion Techniques to Improve Information Quality: Taxonomy, Opportunities and Challenges.” Information Fusion 78: 102–37.
[21] Mussina, Damira, Aidana Irmanova, Prashant K Jamwal, and Mehdi Bagheri. 2020. “Multi-Modal Data Fusion Using Deep Neural Network for Condition Monitoring of High Voltage Insulator.” IEEE Access 8: 184486–96.
[22] Cai, Yi, Binil Starly, Paul Cohen, and Yuan-Shin Lee. 2017. “Sensor Data and Information Fusion to Construct Digital-Twins Virtual Machine Tools for Cyber-Physical Manufacturing.” Procedia Manufacturing 10: 1031–42.
[23] Soleimani, Masoud, Hossein Naderian, Amir Hossein Afshinfar, Zoha Savari, Mahtab Tizhari, Seyed Reza Agha Seyed Hosseini, and others. 2023. “A Method for Predicting Production Costs Based on Data Fusion from Multiple Sources for Industry 4.0: Trends and Applications of Machine Learning Methods.” Computational Intelligence and Neuroscience 2023.
[24] Fortino, Giancarlo, Stefano Galzarano, Raffaele Gravina, and Wenfeng Li. 2015. “A Framework for Collaborative Computing and Multi-Sensor Data Fusion in Body Sensor Networks.” Information Fusion 22: 50–70.
[25] Deng, Zi-Li, Yuan Gao, Lin Mao, Yun Li, and Gang Hao. 2005. “New Approach to Information Fusion Steady-State Kalman Filtering.” Automatica 41 (10): 1695–1707.

Doi: https://doi.org/10.54216/FPA.140217
Received: July 17, 2023 Revised: November 12, 2023 Accepted: January 17, 2024