J. Electrical Systems 20-10s (2024): 7482-7489

[1]Motlatsi Cletus Lehloka

An Application of Closest Pair of Points Algorithm to Detect the Outliers on the Fixed Solar System

Journal of Electrical Systems

Abstract: - Renewable energy resources are regarded as clean energy sources and effective utilization of these resources reduces ecological effects, produces little secondary waste and they are feasible in light of present and future economic and social societal demands. Solar energy is a radiant light and heat from the sun that may be harvested through a variety of methods, including solar power for producing electricity, solar thermal energy, and solar architecture. Solar technologies generate electricity from the sun, which is considered to be a clean energy source. Solar energy technologies offer a tremendous chance for mitigating greenhouse gas emissions and lowering global warming by replacing traditional energy sources. However, the technologies are not exempted from the operational challenges such as the ever-fluctuating ambient conditions. The intention of this study is to present the experimental outcomes of an application of the closest pair of points algorithm to detect outliers on the fixed solar system. The closest pair of points algorithm is used to find outliers in the system based on the measured voltage from the photovoltaic (PV) panel. The algorithm examines the PV panel's immediate voltage and compares it to earlier voltage samples to assess whether there is a significant voltage variation that might turn into outliers. The algorithm proved to be extremely precise and efficient in detecting outliers on the fixed solar system. However, the efficiency of the algorithm should yet be verified for larger PV arrays to determine whether it will withstand the test.

Keywords: Photovoltaic, outliers, closest pair of points algorithm and renewable energy.

I. Introduction

Renewable energy is generated from infinite, naturally replenished resources such as the sun, tides, and wind [1]. Renewable energy may be used to generate power [2]. In contrast, non-renewable energy is derived from finite sources such as coal, natural gas, and oil (fossil fuels) [3]. Fossil fuel burning accounts for the bulk of air pollution produced worldwide [4]. When fossil fuels are burnt, nitrogen oxides infiltrate the atmosphere, contributing to the formation of smog and acid rain [5]. Replacing fossil fuel-based power plants with renewable energy sources such as wind and solar is critical to stabilizing climate change and attaining net zero carbon emissions [6].

One of the cleanest methods for generating electricity is the use of solar energy resources [7]. The most significant advantages of solar technologies are that they do not emit carbon dioxide or harm the atmosphere when used to generate electricity [8]. Greenhouse gases are emitted over the lifetime of some technologies, such as during production, although overall emissions are substantially lower than those from fossil fuels [9]. Solar energy is captured from the sun and converted into electricity, which is then sent into a power system or stored for future use [10]. When sunlight strikes a solar cell, a tiny electric voltage emerges due to the photovoltaic (PV) effect, which occurs between a metal and a semiconductor such as silicon, or two distinct semiconductors. [11]. The PV action releases electrons, which flow across an external circuit because semiconductors have a natural difference in electric potential (voltage) [12].

Thousands of kilowatts of electricity may be produced by building solar panel arrays out of a huge number of solar cells [13]. An average of 342 watts of solar energy are received by each square meter of Earth over the course of a year [14]. The amount of solar radiation that reaches Earth in an hour is more than sufficient to power the globe for the entire year [15]. The black dots in fig. 1 depict places that might create enough energy from sunlight to power the entire world for the year [16]. Despite how much energy the sun gives the globe, there are still certain obstacles for solar systems to overcome. Ensuring the quality of data is crucial for analyzing the performance and dependability of solar energy systems [17].

When it comes to data integrity, incorrect data is a typical concern with PV monitoring systems [18]. Actual infield measurements frequently show inaccurate data (i.e. gaps, missing data, erroneous and outlier values) due to equipment/component failures, power outages, communication issues, or maintenance-related interruptions, all of which can seriously skew the data-based analysis's conclusions [19]. For this reason, before beginning any

[1] University of South Africa, Department of Electrical and Smart Systems Engineering, South Africa

Email: [1]lehlomc@unisa.ac.za

analysis, erroneous data should be found and dealt with accordingly [20]. The purpose of this paper is to present the empirical findings of applying the closest pair of points (CPP) algorithm in order to detect the outliers in a fixed solar system's data for optimization.

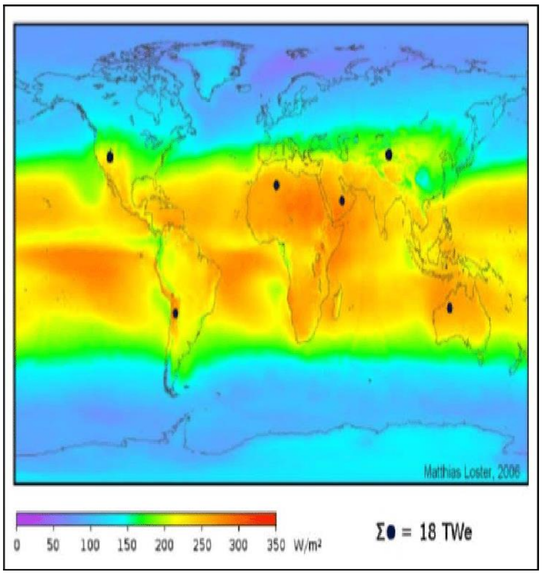

Fig. 1. Annual average solar irradiance distribution over the surface of the Earth [21]

II. LITERATURE REVIEW

Solar power data analysis is a useful tool to utilize when making energy policy decisions since it may assist improve the performance, dependability, and sustainability of solar energy systems [22]. It is necessary to adhere to certain best practices that may guarantee the quality, accuracy, and relevance of the data in order to use it successfully [23]. The initial stage of analysing data on solar power involves gathering pertinent and dependable information from several sources, including weather stations, inverters, meters, sensors, solar panels, and grid operators [24]. Regular checks on data quality and validity, as well as correcting any anomalies such as missing values, outliers, or errors, are critical in PV systems data analysis [25]. It is imperative that the algorithms employed for gathering data are standardized, consistent, and work harmoniously with the tools used for analysis [26].

Determining the readiness of the solar power data analysis involves cleaning, filtering, converting, and aggregating the data in accordance with the objectives and criteria [27]. Appropriate strategies and algorithms for handling data volume, diversity, and velocity have to be considered [28]. It is critical to derive meaningful insights and patterns from acquired data while analyzing solar power [29].

The closest pair of points (CPP) algorithm can be used to detect outliers in a fixed solar system subsequently enhance its performance [30]. The closest pair between the two halves and the minimum of the closest pairings inside each half is called the CPP [31]. The CPP algorithm is a key approach for stiff registration between two-point sets [32]. The algorithm offers several applications ranging from 3D reconstruction to robotics [33]. The algorithm's essential elements include resilience and sensitivity to outliers, missing data, and partial overlaps [34]. The algorithm alternates between finding the closest point in the target set and minimizing the distance between related points, resulting in a locally optimum alignment [35]. However, the algorithm may have delayed convergence owing to its linear convergence rate [28]. The two most popular approaches for detecting outliers are density-based and distance-based approaches (fig. 2) with points found to be outside the predetermined range classified as outliers [36].

Initially, finding outliers was driven by data cleansing, which involved eliminating outliers from the dataset to improve the fit of parametric statistical models to training data [38]. In recent years, there has been an increased focus on outliers themselves, as they frequently provide significant and intriguing information. Examples of such information include network cyberattacks, mechanical issues resulting from faulty industrial machinery, and more [39]. Outliers can be categorized in two categories: one (1) point outlier and two (2) collective outliers, depending on how many data examples make up a deviant pattern [40]. A single data point that significantly differs from the rest of the dataset is called a point outlier [41]. Collective outliers are data points that stand out from the rest of the dataset [42].

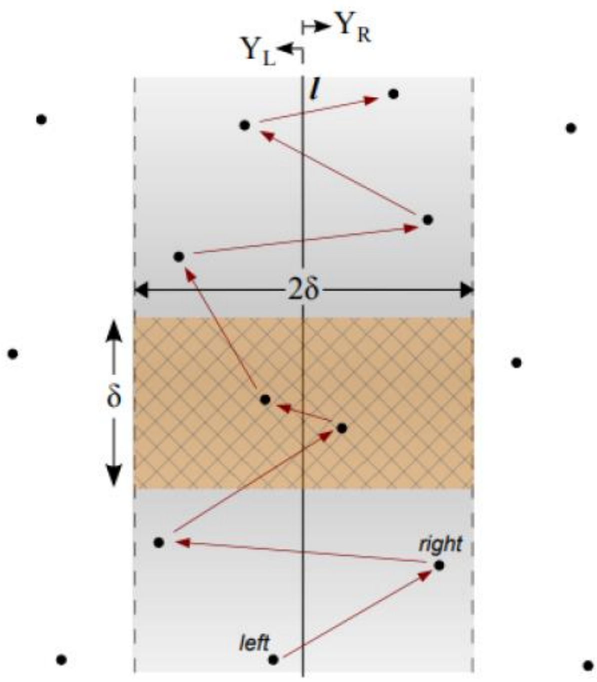

Fig. 2. Closest pair of points [37]

Outlier identification is currently one of the primary responsibilities of time series data mining and has drawn interest from several scholars and practitioners [43]. Abnormal behaviours throughout time are examined through the examination of outliers in time series data [44]. Outliers in time series might have two alternative interpretations (fig. 3), depending on the analyst's interest or the circumstance under consideration [45]. These observations are connected to noisy, erroneous, or undesirable data, which are not of interest to the analyst [46]. To improve the quality of the data and generate a cleaner dataset that can be utilized by other data mining algorithms, outliers in these situations should be eliminated or rectified [47].

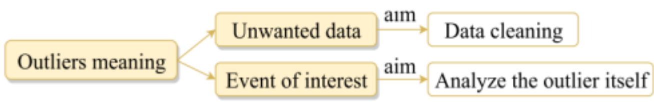

Fig. 3. Meaning of the outliers in time series data depending on the aim of the analyst [48]

The paper aims to provide the experimental results of the closest pair of points algorithm used for detecting outliers on the fixed solar system for optimization purposes.

III. Research methodology

A 310 W YL310P-35b polycrystalline PV panel was used in this study. It was utilized since it is less expensive and performs better in locations with direct sunlight [49]. LabVIEW program was used to measure and analyze data from the PV panel. The system was erected on the top of the Euclid building at the University of South Africa (UNISA) science campus in Florida. The algorithm was programmed using LabVIEW software.

The algorithm sorts points in accordance with their X-Y coordinates. It then divides the sorted set into half, making equal-sized. Recursively, it solves the problem in the left and right subsets. This provides the minimal distances on the left and right, respectively. Then, the algorithm establishes the shortest distance between two places that are on opposite sides of the vertical line. The final solution is the minimal distance between the point pairs. The points that fall beyond the established distance are considered to be outliers.

IV. RESULTS AND DISCUSSION

The basic principle is that an outlier point may be located in a sparse zone, whereas normal points can be found in a denser region. Table 1 lists the PV panel voltage readings for each time slot that was subsequently used to establish fig. 4. From table 1, it can be noted that immediately at 7 am, there is a voltage-shooting of 67,48 V that is even greater than the maximum reading at 12 noon. It is worth noting that this is an alarming reading as compared to table 2 (that was also used to establish fig. 5). At the same period (7 am) the system's voltage is 10,06 V which is reasonable looking at the time of the day, resulting with the leaner-graph. The same behaviour (voltage shooting/drop) repeats again 2 pm and 5pm.

TABLE I: Hourly voltage readings with outliers

Time (Hrs)	**Voltage (V)**
06:00	0,2
07:00	64,78
08:00	23,34
09:00	36,24
10:00	48,37
11:00	58,27
12:00	63,77
13:00	61,97
14:00	4,6
15:00	42,73
16:00	30,21
17:00	66,11
16:00	3,83

TABLE II: Hourly voltage readings without outliers

Time (Hrs)	**Voltage (V)**
06:00	0,2
07:00	10,06
08:00	23,34
09:00	36,24
10:00	48,37
11:00	58,27
12:00	63,77
13:00	61,97
14:00	53,9
15:00	42,73

16:00	31,21
17:00	17,08
16:00	3,83

Fig. 4 depicts collective outlier points that are significantly apart from the rest of the densely packed clusters and hence are outlier points in the dataset, while fig. 5 depicts the data set without outliers after the CPP algorithm intervention. The fundamental principle of outlier point detection is that normal points are found in higher density populations, but outlier points persist in sparse populations. The basic presumption of the closest pair of points approach is that the distances between outlier points are wider than those between normal data points. It is evident from fig.5 that the algorithm is reliable as far as handling outliers in the dataset is concerned. The data set without outliers after the CPP algorithm intervention is shown in fig. 5.

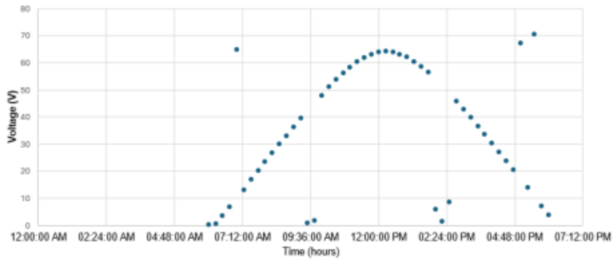

Fig. 4. Voltage readings with outliers

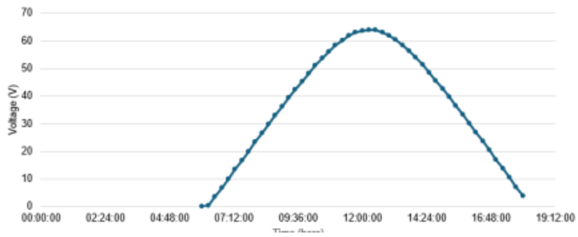

Fig. 5. Voltage readings without outliers.

V. Conclusion

This paper presented the results of CPP algorithm in approach for outlier detection. The significance of the algorithm was outlined together with the effects of the outliers in data mining. It was discovered that the algorithm does not rely on underlying distribution of data to detect outliers and it is easy to implement. Subsequently, the algorithm detects and eliminates outliers vigorously. As much as the algorithm was discovered to be robust in dealing with outliers, it is critical to test it on large-scale PV panels.

Acknowledgment

I would like to express my gratitude to Prof. Z. Wang for guidance and The University of South Africa for the financial support.

References

[1] S. Dey, A. Sreenivasulu, G. Veerendra, K. V. Rao, and P. A. Babu, “Renewable energy present status and future potentials in india: An overview,” Innovation and Green Development, vol. 1, no. 1, p. 100006, 2022.

[2] T.-Z. Ang, M. Salem, M. Kamarol, H. S. Das, M. A. Nazari, and N. Prabaharan, “A comprehensive study of renewable energy sources: Classifications, challenges and suggestions,” Energy Strategy Reviews, vol. 43, p. 100939, 2022.

[3] M. Amin, H. H. Shah, A. G. Fareed, W. U. Khan, E. Chung, A. Zia, Z. U. Rahman Farooqi, and C. Lee, “Hydrogen production through renewable and non-renewable energy processes and their impact on climate change,” International Journal of Hydrogen Energy, vol. 47, no. 77, pp. 33112–33134, 2022.

[4] Jiang, R. Xing, Z. Luo, W. Huang, F. Yi, Y. Men, N. Zhao, Z. Chang, J. Zhao, B. Pan, and G. Shen, “Pollutant emissions from biomass burning: A review on emission characteristics, environmental impacts, and research perspectives,” Particuology, vol. 85, pp. 296–309, 2024.

[5] A. K. Sharma, M. Sharma, A. K. Sharma, M. Sharma, and M. Sharma, “Mapping the impact of environmental pollutants on human health and environment: A systematic review and meta-analysis,” Journal of Geochemical Exploration, vol. 255, p. 107325, 2023.

[6] Y. Zhou, “Worldwide carbon neutrality transition? energy efficiency, renewable, carbon trading and advanced energy policies,” Energy Reviews, vol. 2, no. 2, p. 100026, 2023.

[7] A. Shamoon, A. Haleem, S. Bahl, M. Javaid, S. Bala Garg, R. Chandmal Sharma, and J. Garg, “Environmental impact of energy production and extraction of materials - a review,” Materials Today: Proceedings, vol. 57, pp. 936–941, 2022. Third International Conference on Aspects of Materials Science and Engineering.

[8] T. M. Gur, “Carbon dioxide emissions, capture, storage and utilization: ¨ Review of materials, processes and technologies,” Progress in Energy and Combustion Science, vol. 89, p. 100965, 2022.

[9] P. T. Aakko-Saksa, K. Lehtoranta, N. Kuittinen, A. Jarvinen, J.-P. ¨ Jalkanen, K. Johnson, H. Jung, L. Ntziachristos, S. Gagne, C. Takahashi, ´ P. Karjalainen, T. Ronkko, and H. Timonen, “Reduction in greenhouse ¨ gas and other emissions from ship engines: Current trends and future options,” Progress in Energy and Combustion Science, vol. 94, p. 101055, 2023.

[10] M. Soroush and Y. Hajimolana, “Sunlight harvesting,” Computers Chemical Engineering, vol. 170, p. 108103, 2023.

[11] M. A. Hossain, M. Z. Haque, S. Reza, B. Khan, M. N. Islam, et al., “A review of renewable energy in system bangladesh and performance analysis of mini hybrid energy generation model to cope with the system,” 2022.

[12] L. Xu, Y. Yang, Y. Mao, and Z. Li, “Self-powerbility in electrical stimulation drug delivery system,” Advanced Materials Technologies, vol. 7, no. 2, p. 2100055, 2022.

[13] J. Tang, H. Ni, R.-L. Peng, N. Wang, and L. Zuo, “A review on energy conversion using hybrid photovoltaic and thermoelectric systems,” Journal of Power Sources, vol. 562, p. 232785, 2023.

[14] T. Thennakoon, D. Weerarathna, H. Hewage, W. Jayathma, D. Gamage, L. Sulaksha, A. Alahakoon, P. Senarath, and G. Bandara, “Overview of solar electricity in sri lanka and recycling processes,”

[15] C. M. S. Kumar, S. Singh, M. K. Gupta, Y. M. Nimdeo, R. Raushan, A. V. Deorankar, T. A. Kumar, P. K. Rout, C. Chanotiya, V. D. Pakhale, et al., “Solar energy: A promising renewable source for meeting energy demand in indian agriculture applications,” Sustainable Energy Technologies and Assessments, vol. 55, p. 102905, 2023.

[16] B. Iglinski, M. Skrzatek, W. Kujawski, M. Cichosz, and R. Buczkowski, ´ “Swot analysis of renewable energy sector in mazowieckie voivodeship (poland): current progress, prospects and policy implications,” Environment, Development and Sustainability, vol. 24, no. 1, pp. 77–111, 2022.

[17] O. A. Alimi, E. L. Meyer, and O. I. Olayiwola, “Solar photovoltaic modules’ performance reliability and degradation analysis—a review,” Energies, vol. 15, no. 16, p. 5964, 2022.

[18] M. A. Husnoo, A. Anwar, N. Hosseinzadeh, S. N. Islam, A. N. Mahmood, and R. Doss, “False data injection threats in active distribution systems: A comprehensive survey,” Future Generation Computer Systems, vol. 140, pp. 344–364, 2023.

[19] Y. Zhuang, W. Chen, T. Jin, B. Chen, H. Zhang, and W. Zhang, “A review of computer vision-based structural deformation monitoring in field environments,” Sensors, vol. 22, no. 10, p. 3789, 2022.

[20] H. T. Reda, A. Anwar, and A. Mahmood, "Comprehensive survey and taxonomies of false data injection attacks in smart grids: attack models, targets, and impacts," Renewable and Sustainable Energy Reviews, vol. 163, p. 112423, 2022.

[21] F. M. Guangul and G. T. Chala, "Solar energy as renewable energy source: Swot analysis," 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), pp. 1–5, 2019.

[22] T. Ahmad, R. Madonski, D. Zhang, C. Huang, and A. Mujeeb, "Datadriven probabilistic machine learning in sustainable smart energy/smart energy systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm," Renewable and Sustainable Energy Reviews, vol. 160, p. 112128, 2022.

[23] A. S. Albahri, A. M. Duhaim, M. A. Fadhel, A. Alnoor, N. S. Baqer, L. Alzubaidi, O. S. Albahri, A. H. Alamoodi, J. Bai, A. Salhi, et al., "A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion," Information Fusion, 2023.

[24] J. Li, M. S. Herdem, J. Nathwani, and J. Z. Wen, "Methods and applications for artificial intelligence, big data, internet of things, and blockchain in smart energy management," Energy and AI, vol. 11, p. 100208, 2023.

[25] V. Balaram and M. Satyanarayanan, "Data quality in geochemical elemental and isotopic analysis," Minerals, vol. 12, no. 8, p. 999, 2022.

[26] J. P. Castellanos Ardila, B. Gallina, and F. Ul Muram, "Compliance checking of software processes: A systematic literature review," Journal of Software: Evolution and Process, vol. 34, no. 5, p. e2440, 2022.

[27] V. Harish, N. Anwer, and A. Kumar, "Applications, planning and sociotechno-economic analysis of distributed energy systems for rural electrification in india and other countries: A review," Sustainable Energy Technologies and Assessments, vol. 52, p. 102032, 2022.

[28] A. G. Gad, "Particle swarm optimization algorithm and its applications: a systematic review," Archives of computational methods in engineering, vol. 29, no. 5, pp. 2531–2561, 2022.

[29] H. Liao, E. Michalenko, and S. C. Vegunta, "Review of big data analytics for smart electrical energy systems," Energies, vol. 16, no. 8, p. 3581, 2023.

[30] M. Ratz, P. Henkel, P. Stoffel, R. Streblow, and D. Muller, "Identifying ¨ the validity domain of machine learning models in building energy systems," Energy and AI, p. 100324, 2023.

[31] J. Madake, S. Bhatlawande, S. Shilaskar, S. Kulkarni, K. Mehetre, N. Wani, and S. Bangadkar, "Point cloud registration for 3d modelling," in 2022 3rd International Conference for Emerging Technology (INCET), pp. 1–8, IEEE, 2022.

[32] X. Yuan and A. Maharjan, "Non-rigid point set registration: recent trends and challenges," Artificial Intelligence Review, vol. 56, no. 6, pp. 4859– 4891, 2023.

[33] H. Ding, B. Zhang, J. Zhou, Y. Yan, G. Tian, and B. Gu, "Recent developments and applications of simultaneous localization and mapping in agriculture," Journal of field robotics, vol. 39, no. 6, pp. 956–983, 2022.

[34] S. C. K. Tekouabou, E. B. Diop, R. Azmi, R. Jaligot, and J. Chenal, "Reviewing the application of machine learning methods to model urban form indicators in planning decision support systems: Potential, issues and challenges," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 8, pp. 5943–5967, 2022.

[35] M. Chaudhry, I. Shafi, M. Mahnoor, D. L. R. Vargas, E. B. Thompson, and I. Ashraf, "A systematic literature review on identifying patterns using unsupervised clustering algorithms: A data mining perspective," Symmetry, vol. 15, no. 9, p. 1679, 2023.

[36] M. N. K. Sikder and F. A. Batarseh, "Outlier detection using ai: a survey," AI Assurance, pp. 231–291, 2023.

[37] J. C. Pereira and F. G. Lobo, “An optimized divide-and-conquer algorithm for the closest-pair problem in the planar case,” Journal of Computer Science and Technology, vol. 27, no. 4, pp. 891–986, 2012.

[38] K. Mallikharjuna Rao, G. Saikrishna, and K. Supriya, “Data preprocessing techniques: emergence and selection towards machine learning models-a practical review using hpa dataset,” Multimedia Tools and Applications, vol. 82, no. 24, pp. 37177–37196, 2023.

[39] M. H. Rahman, T. Wuest, and M. Shafae, “Manufacturing cybersecurity threat attributes and countermeasures: Review, meta-taxonomy, and use cases of cyberattack taxonomies,” Journal of Manufacturing Systems, vol. 68, pp. 196–208, 2023.

[40] J. Rahnenfuhrer, R. De Bin, A. Benner, F. Ambrogi, L. Lusa, A.- ¨ L. Boulesteix, E. Migliavacca, H. Binder, S. Michiels, W. Sauerbrei, et al., “Statistical analysis of high-dimensional biomedical data: a gentle introduction to analytical goals, common approaches and challenges,” BMC medicine, vol. 21, no. 1, p. 182, 2023.

[41] B. Dastjerdy, A. Saeidi, and S. Heidarzadeh, “Review of applicable outlier detection methods to treat geomechanical data,” Geotechnics, vol. 3, no. 2, pp. 375–396, 2023.

[42] H. Alimohammadi and S. N. Chen, “Performance evaluation of outlier detection techniques in production timeseries: A systematic review and meta-analysis,” Expert Systems with Applications, vol. 191, p. 116371, 2022.

[43] J. S. Sunny, C. P. K. Patro, K. Karnani, S. C. Pingle, F. Lin, M. Anekoji, L. D. Jones, S. Kesari, and S. Ashili, “Anomaly detection framework for wearables data: A perspective review on data concepts, data analysis algorithms and prospects,” Sensors, vol. 22, no. 3, p. 756, 2022.

[44] W. Yu and Q. Huang, “A deep encoder-decoder network for anomaly detection in driving trajectory behavior under spatio-temporal context,” International Journal of Applied Earth Observation and Geoinformation, vol. 115, p. 103115, 2022.

[45] C. V. Ribeiro, A. Paes, and D. de Oliveira, “Ais-based maritime anomaly traffic detection: A review,” Expert Systems with Applications, p. 120561, 2023.

[46] M. M. Ahmed, M. N. Khan, A. Das, and S. E. Dadvar, “Global lessons learned from naturalistic driving studies to advance traffic safety and operation research: A systematic review,” Accident Analysis & Prevention, vol. 167, p. 106568, 2022.

[47] P. Saraswat and S. Raj, “Data pre-processing techniques in data mining: A review,” International Journal of Innovative Research in Computer Science & Technology, vol. 10, no. 1, pp. 122–125, 2022.

[48] A. Blazquez-Garcıa, A. Conde, U. Mori, and J. A. Lozano, “A review on outlier/anomaly detection in time series data,” ACM Computing Surveys (CSUR), vol. 54, no. 3, pp. 1–33, 2021.

[49] M. C. Lehloka, J. A. Swart, and P. E. Hertzog, “Linear regression algorithm results for a pv dual-axis tracking-type system,” International Journal of Electrical and Electronic Engineering & Telecommunications, vol. 10, no. 2, pp. 139–144, 2021.