Unveiling the Operational Patterns of Global LNG Terminal Points: A Multi-algorithmic Clustering Approach

Haoda Wen

Internet of Things Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
WHD@bupt.edu.cn

Abstract. This study embarked on a mission to solve a complex issue in the field of maritime logistics - understanding and classifying the operational patterns of Liquified Natural Gas (LNG) terminal points across the globe. Given the significant role of LNG shipping in global energy transport, identifying these patterns holds substantial value for optimizing shipping operations and improving safety. The adopted methodology was a novel multi-algorithmic approach using machine learning, specifically leveraging the BIRCH, KMeans, and DBSCAN clustering algorithms. These tools enabled this study to scrutinize a comprehensive dataset, capturing the operational dynamics of LNG terminal points based on draft depth alterations. This approach facilitated a more nuanced and sophisticated understanding of the terminal points, unlike conventional methodologies. The research findings delineated a detailed and holistic geospatial distribution of LNG terminal points. The results were testament to the effectiveness of the proposed approach as the generated cluster graphs closely mirrored actual site maps, illustrating a high degree of precision and robustness. It was demonstrated that the operational patterns derived from this methodology can provide superior insights for logistics planning, potentially allowing for more efficient and safer maritime operations. However, it is important to acknowledge that the choice of clustering algorithm can influence the resolution and accuracy of the clustering results, indicating the possibility for further refinement. Moving forward, this work establishes a promising foundation for optimizing LNG logistics and developing predictive traffic models, with potential applications extending to other domains within maritime data analysis.

Keywords: Machine Learning, Clustering Algorithm, LNG Terminal Points

1 Introduction

Liquid Natural Gas (LNG) forms a critical part of the global energy mix, with shipping serving as the primary mode of its long-distance transport. These LNG ships, with their unique operations and patterns, constitute a complex network that holds

P. Kar et al. (eds.), *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)*, Advances in Computer Science Research 108,
https://doi.org/10.2991/978-94-6463-300-9_28

significant influence over the international energy markets and geopolitics. A deeper understanding of these patterns would provide valuable insights to stakeholders in the maritime industry and beyond.

Specifically, in the context of LNG shipping, previous research mainly concentrated on aspects such as market analysis [1, 2], logistical issues [3, 4], and emission impacts [5, 6], largely overlooking the operational patterns at LNG terminal points. The few studies that attempted to unravel these patterns often relied on traditional statistical methods, which, while insightful, might not capture the complexity and dynamics of such patterns effectively.

Despite the critical role of LNG terminal points in shaping the global LNG shipping network, there is a conspicuous gap in the literature concerning the data-driven identification and classification of these points. Moreover, traditional methods often require explicit hypothesis formulation and are based on predefined notions about the underlying patterns. On the contrary, machine learning techniques, such as the clustering algorithms can learn directly from the data, capturing patterns that might be missed by other methods [7-9].

This research, therefore, is motivated by the necessity to fill this gap, offering a comprehensive, data-driven analysis of LNG terminal points. This study goes beyond the scope of the previous research by leveraging machine learning techniques to uncover operational patterns at these terminal points from an extensive dataset. It thereby contributes to a deeper understanding of the global LNG shipping network and serves as a novel precedent in employing advanced data science techniques for this purpose. This approach not only supplements the existing knowledge base but also sets the stage for future research in this domain.

Recognizing the gap in the current literature, this study takes an innovative and data-driven approach to explore the operational patterns at LNG terminal points, namely berthing, inbound, and outbound points. Leveraging a comprehensive dataset of LNG ship movements and draught (water depth) details, the research involves a combination of machine learning techniques, specifically the clustering algorithms BIRCH, KMeans, and DBSCAN, to analyze and classify these terminal points. To solve the limitation mentioned above, this paper focuses on the simultaneous completion of medical images classification and denoising within the same end-to-end model. By consolidating the two tasks together, the model leverages shared knowledge and complementary information, leading to possibly improved performance and efficiency. In this paper, a model based on an autoencoder with an auxiliary classifier is applied. Specifically, the encoder of model first compressed image information into a bottleneck layer with CNN. Then based on the compressed representation in the bottleneck layer, an auxiliary classifier utilizes a feed-forward network for classification, while the decoder employs a multi-layer CNN architecture for denoising purposes. By jointly training and optimizing both objectives, the model can simultaneously achieve classification and denoising of given degraded images.

2 Method

An overview of the work shows the meticulous steps taken in pre-processing the data and optimizing the algorithms' parameters for robust results. After data cleaning and partitioning, the methodology employs a three-pronged approach, using the three clustering algorithms independently to analyze the data. This approach provides an enriched perspective and more robust findings, as it enables cross-comparison of the results. The general working flow chart is as Fig. 1.

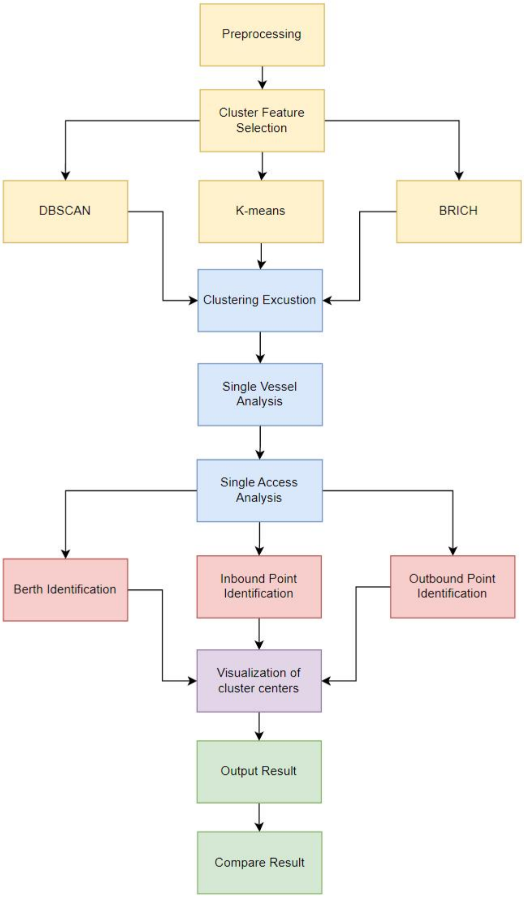

Fig. 1. Workflow chart of this study (Photo/Picture credit: Original).

2.1 Data Collection and Preprocessing

The data used for this study was derived from an expansive data set comprising of 3,590,578 data points collected from LNG ships over a period of three months. The data was obtained while the ship's speed was stationary (less than 1 knot). The variables included in this data set are: Maritime Mobile Service Identity (mmsi), Unix timestamp (in seconds), navigational status, speed, longitude, latitude, and draft.

Initial data processing involved loading the data set and conducting necessary preprocessing to refine it. This process encompassed the removal of rows containing null values, as well as entries where the draft was recorded as zero. This was done to ensure that the data set was accurate, comprehensive, and relevant for the objectives of this research.

2.2 Cluster Method

Following data preprocessing, the selection of clustering features was focused. The geographic coordinates of longitude and latitude were identified as the primary features for clustering in this research, aiming to determine the geographical positions of the ships.

2.2.1 Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH)

BIRCH is a hierarchical clustering method that is specially designed for large datasets. The algorithm operates without having to store all data in memory, thus making it efficient for handling the large dataset used in this study. BIRCH incrementally and dynamically clusters incoming multi-dimensional metric data points to try to produce the best quality clustering. The schematic of the BIRCH is shown in Fig. 2.

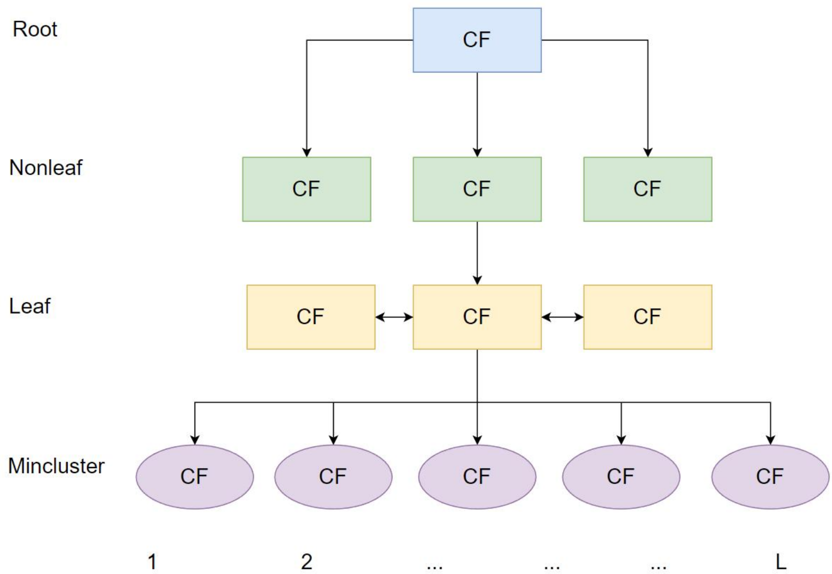

Fig. 2. CF tree of BRICH (Photo/Picture credit: Original).

2.2.2 K-Means

K-Means shown in Fig. 3 is a partitioning clustering technique that divides the data into non-overlapping subsets or clusters without any cluster internal structure. The number of clusters (K) is user-defined. It uses the distance between points to partition them into clusters aiming to minimize the intra-cluster distances.

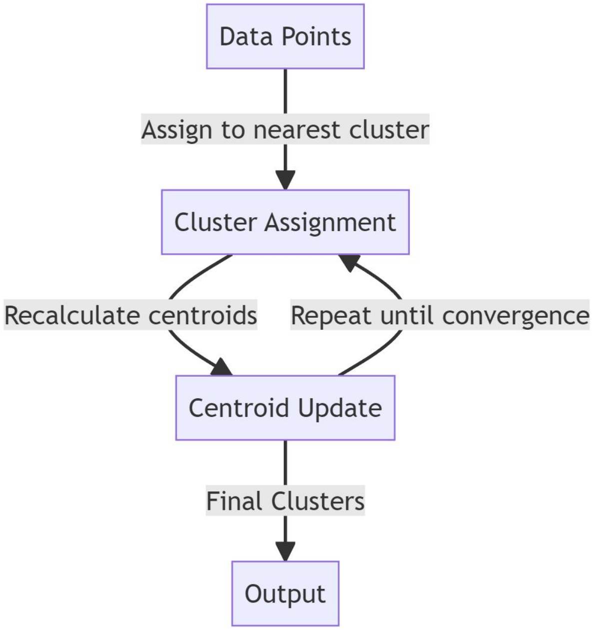

Fig. 3. Steps of K-Means (Photo/Picture credit: Original).

2.2.3 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

In contrast to K-Means, obviates the need for predefining the number of clusters in advance. DBSCAN shown in Fig. 4. is a density-based clustering algorithm: given a set of points in a space, it groups together points that are close to each other based on a distance measurement (usually Euclidean distance) and a minimum number of points. It is also capable of identifying and designating as outliers those points that reside within regions of low data density.

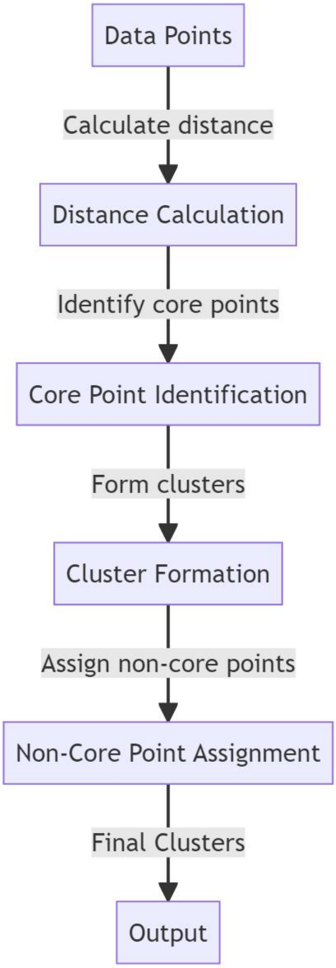

Fig. 4. Steps of DBSCAN (Photo/Picture credit: Original).

By leveraging multiple clustering algorithms, the strengths of each can be harnessed. This allowed for a rigorous comparison and cross-validation of clustering results, ultimately leading to a more robust and reliable identification of key berthing points. The use of multiple algorithms not only provided an initial estimate of the number of berths but also afforded the flexibility to refine the results for optimal accuracy.

2.3 Site Type Determination

The proposed approach to identifying the nature of berthing points – specifically distinguishing between berths, inbound points, and outbound points – integrated the examination of the draft depth alterations for each vessel within each cluster. This study adopted this method based on the understanding of the typical activities that occur at these points.

In instances of Berths, this study anticipated a relative stability in the vessel's draft depth throughout its period of stay. This assumption aligns with the usual practices at berthing points where vessels stay stationary with minimal changes in their cargo, leading to minor fluctuations in the draft depth. To operationalize this, this study marked berths as points where the difference between the maximum and minimum draft depths fell below a predefined threshold.

Inbound Points were characterized by a noticeable augmentation in the vessel's draft depth during its stationing. Specifically, the occurrence of the maximum draft depth posts the minimum draft depth was indicative of an inbound point. This increase in draft depth results from the loading of cargo onto the vessel, which increases its weight and correspondingly, its draft depth.On the other hand, Outbound Points demonstrated a substantial reduction in the vessel's draft depth during its stay. In these cases, the maximum draft depth was recorded before the minimum, indicating the unloading of cargo that decreases the vessel's weight and draft depth.It is essential to note that the proposed methodology encompassed every visit of each vessel to every berthing point. Given that a single vessel could visit the same point multiple times, possibly for different activities (loading, unloading, or remaining stationary), draft depth variations per visit was calculated. This allowed this study to identify the nature of the point based on the changes in draft depth during each visit.

The procedure was carried out as follows: this study began by grouping the data for each cluster using the Maritime Mobile Service Identity (MMSI), the unique identifier of each vessel. Subsequently, the data for each vessel was sorted based on the Unix timestamp and differences between adjacent timestamps were computed. When these differences exceeded a period of 30 days, it was interpreted as the initiation of a new visit. For each visit, the difference between the maximum and minimum draft values was calculated, assigned a positive or negative value based on the chronological sequence, and derived an average for each cluster. This average was utilized to categorize the cluster as either a berth, inbound, or outbound point. Furthermore, the central point of each cluster was computed and added to a list of central coordinates. Finally, these coordinates, along with the classified types, were used to create comprehensive scatter plots for visualization.

3 Results and Discussion

The application of BIRCH, K-Means, and DBSCAN clustering algorithms to the LNG vessel data led to the successful identification of potential LNG sites shown in Fig. 5, Fig.6 and Fig. 7. However, each algorithm presented nuanced advantages and intricacies in their clustering results, reflecting the specificity of their underlying mechanisms.

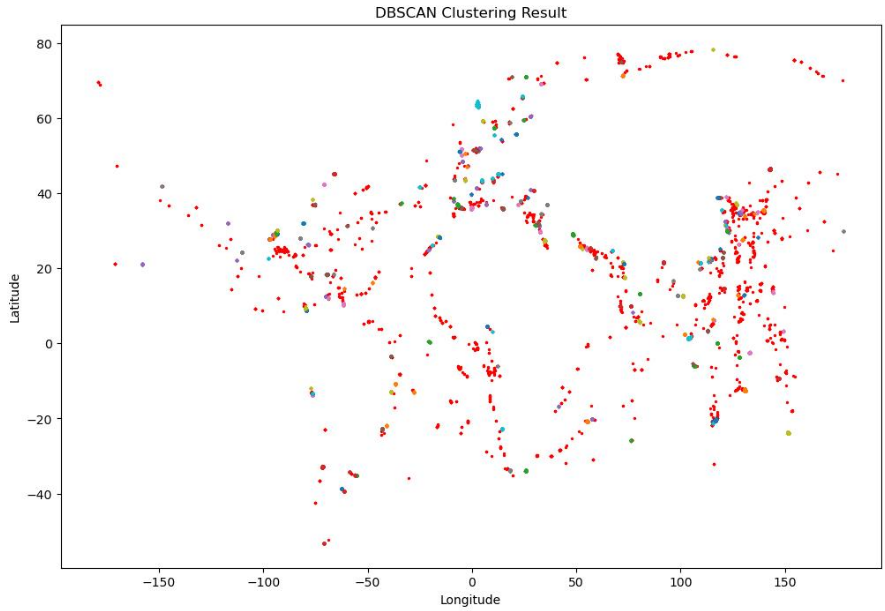

Fig. 5. DBSCAN cluster result (Photo/Picture credit: Original).

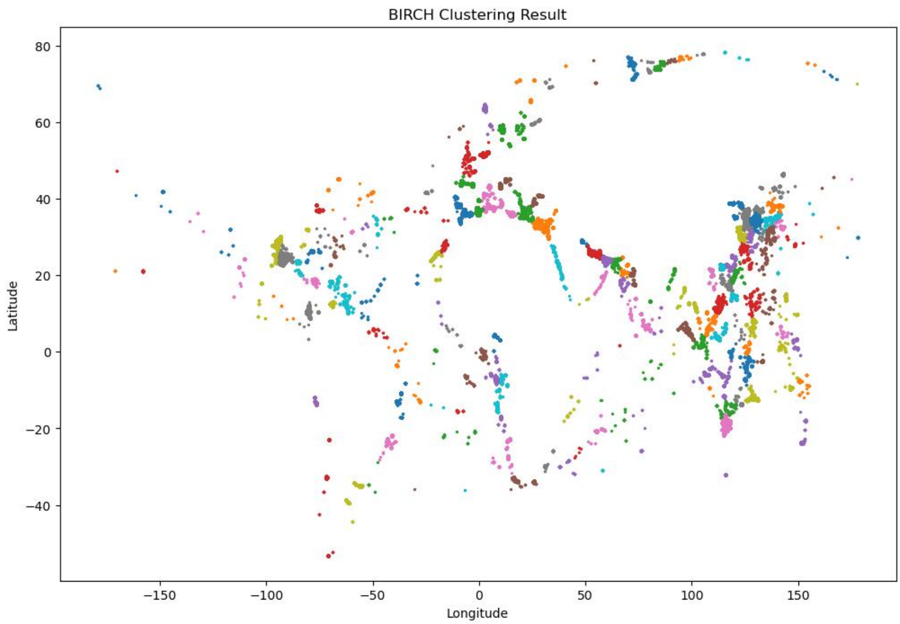

Fig. 6. BRICH cluster result (Photo/Picture credit: Original).

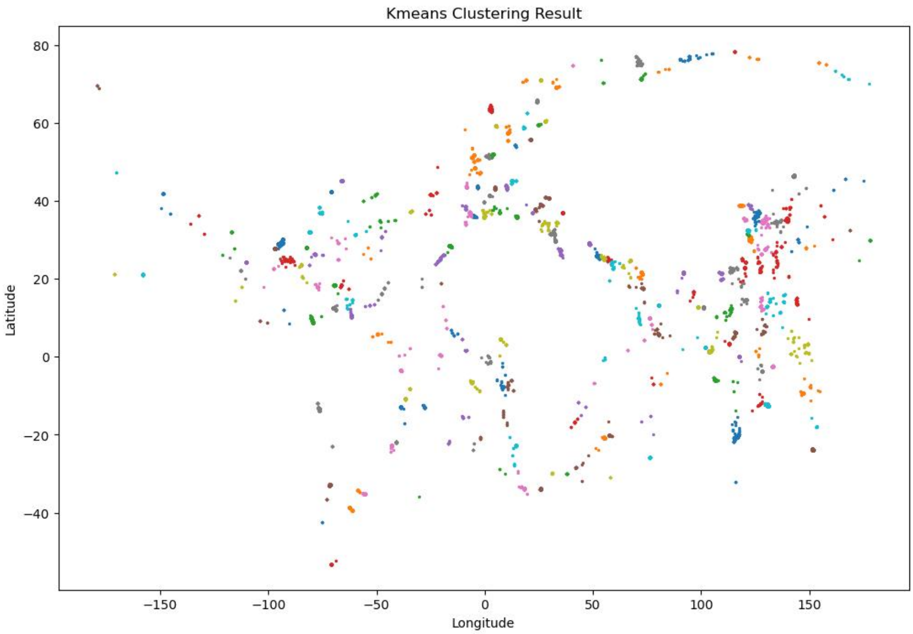

Fig. 7. K-means cluster result (Photo/Picture credit: Original).

The DBSCAN algorithm was particularly adept at handling regions with high site density shown in Fig. 8. It was able to distinguish clusters more accurately in such areas, resulting in a clustering output that closely mirrored reality. This ability proved essential for gaining a preliminary understanding of the overall number of sites, a key consideration in subsequent analysis stages.

On the other hand, BIRCH and K-Means algorithms showcased their strength in generating visually more appealing and relatively discrete clusters shown in Fig.9 and Fig. 10. This aspect of their performance was pivotal in enhancing the recognition of berths, the sites where vessels maintain their position over time. Although both algorithms performed commendably, they offered unique strengths. BIRCH was effective in generating initial clusters, providing a broad overview of the data distribution. K-Means, with its specific focus on partitioning the dataset into distinct groups, yielded a more precise approximation of the stationary sites.

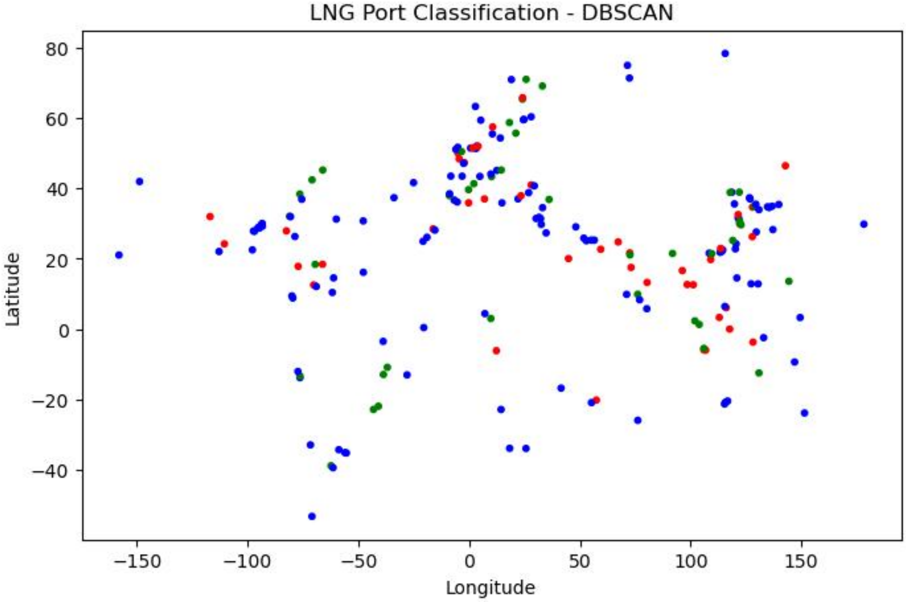

Fig. 8. DBSCAN analysis result. berth: blue, inbound: green, outbound: red (Photo/Picture credit: Original).

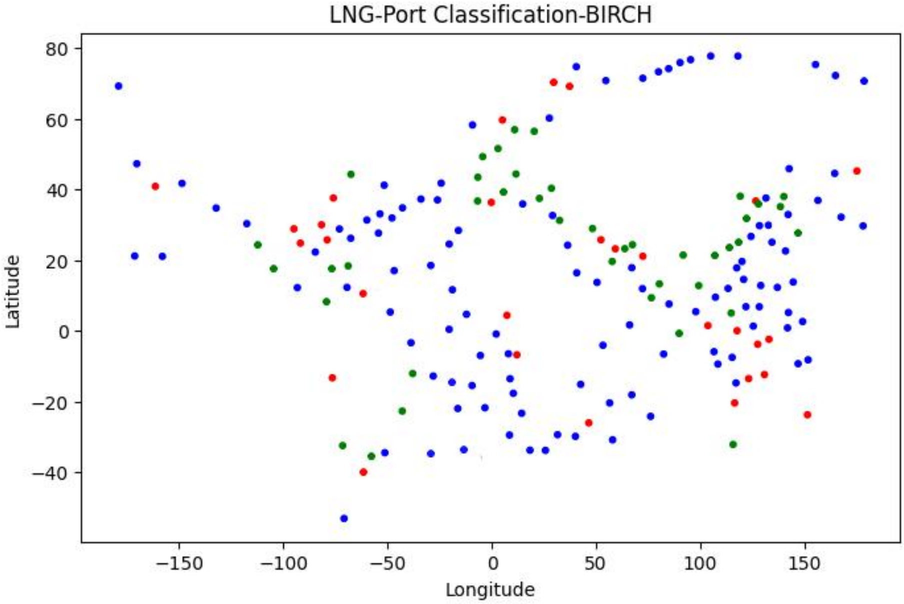

Fig. 9. BRITCH analysis result. berth: blue, inbound: green, outbound: red (Photo/Picture credit: Original).

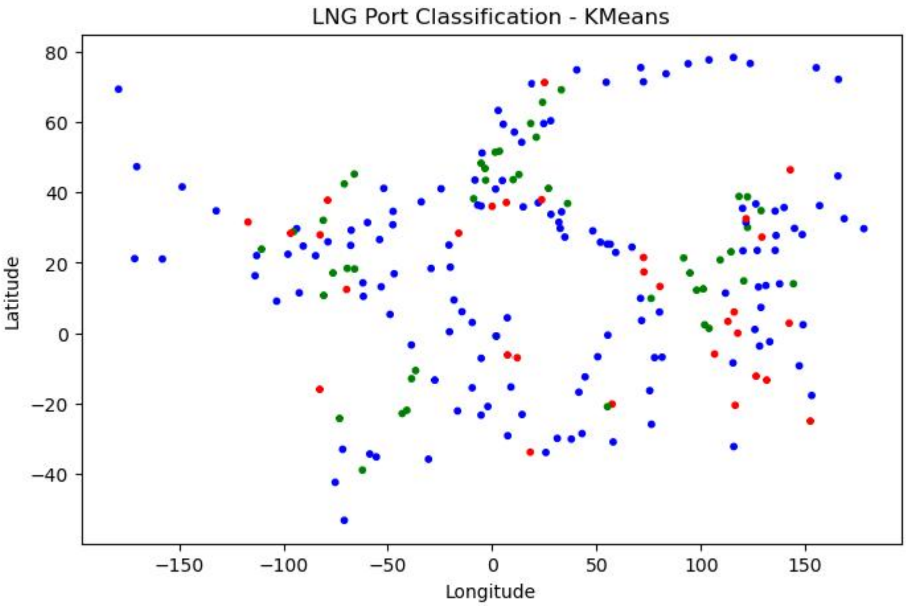

Fig. 10. K-means analysis result. berth: blue, inbound: green, outbound: red (Photo/Picture credit: Original).

The study was further compared with existing documented global sites [10] shown in Fig. 11, acknowledging that some of these sites may differ temporally from the dataset employed in this study. Through this comparative analysis, it was revealed that each of the three employed clustering methods possesses distinctive characteristics, collectively resulting in higher overall accuracy.

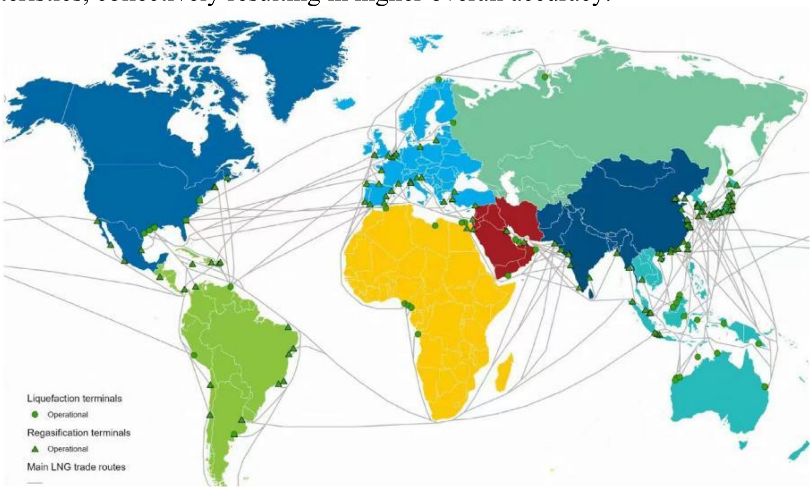

Fig. 11. Major LNG Shipping Routes, 2021 [11].

Collectively, the considered clustering algorithms provided a comprehensive and reliable delineation of potential LNG sites. Their combined implementation offered a higher accuracy rate, contributing to a more authentic and holistic portrayal of the geospatial distribution of the LNG vessels. The identification and classification of LNG sites provide an invaluable perspective on LNG vessels' operations and traffic patterns. Moreover, the distinction between berths, inbound points, and outbound points elucidates the logistics involved in the transportation of LNG.

In contrast to previous studies, this research distinguishes itself through its methodical and comprehensive approach to clustering and site identification. The incorporation of multiple clustering algorithms not only accounts for the intricate geospatial attributes of the data but also enhances the reliability and robustness of the obtained results.

The comparison of the generated cluster graphs with the actual site maps serves as a testament to the accuracy of the clustering methods. This comparison shows significant similarities in terms of the site number and spatial distribution, thus indicating a high level of precision in the research methodology used in this study. However, it is important to acknowledge that the selection of the clustering algorithm can influence the resolution of the clusters and, consequently, impact the identification of the sites.

4 Conclusion

This research explored the operational patterns at LNG terminal points, deploying BIRCH, K-Means, and DBSCAN clustering algorithms to illuminate the complexities of berthing, inbound, and outbound points in the global LNG network. The methodology employed allowed this study to categorize terminal points based on draft depth alterations, yielding a sophisticated picture of the global LNG terminal landscape. The study's findings reveal the sophisticated landscape of LNG terminal points, yielding a more authentic and holistic view of the geospatial distribution of the LNG vessels. The proposed approach presented a novel and superior method for clustering and site identification, exhibiting a high level of precision and robustness.

Notwithstanding, it is imperative to acknowledge that the clustering resolution and site identification can be influenced by the selection of the clustering algorithm, indicating the necessity for additional refinement and optimization of algorithmic parameters. Moving forward, this research lays a foundation for optimizing the logistics of Liquified Natural Gas (LNG) and advancing the development of predictive traffic models. The potential of this approach extends beyond the realm of LNG data, offering the prospect of uncovering novel insights in other maritime domains, ultimately enhancing efficiency and safety in maritime transportation.

References

1. Lee, J. S., Lee, H. T., Cho, I. S.: Maritime traffic route detection framework based on statistical density analysis from AIS data using a clustering algorithm. IEEE Access, 10: 23355-23366 (2022).
2. Zhu, J., Gao, M., Zhang, A., et al.: Multi-Ship Encounter Situation Identification and Analysis Based on AIS Data and Graph Complex Network Theory. Journal of Marine Science and Engineering, 10(10): 1536 (2022).
3. Qi, J., Wang, S.: LNG Bunkering Station Deployment Problem—A Case Study of a Chinese Container Shipping Network. Mathematics, 11, 813 (2023).
4. Li, J., Jiao, H., Wang, J., et al.: Online real-time trajectory analysis based on adaptive time interval clustering algorithm. Big Data Mining and Analytics, 3(2): 131-142 (2020).
5. Ester, M., et al.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Knowledge Discovery and Data Mining (1996).
6. Murray, B., Perera, L. P.: Ship behavior prediction via trajectory extraction-based clustering for maritime situation awareness. Journal of Ocean Engineering and Science, 7(1): 1-13 (2022).
7. Al-Omary, A. Y., Jamil, M. S.: A new approach of clustering based machine-learning algorithm. Knowledge-Based Systems, 19(4): 248-258 (2006).
8. Javidan, S. M., Banakar, A., Vakilian, K. A., et al.: Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning. Smart Agricultural Technology, 3: 100081 (2023).
9. Yu, Q., Chen, P., Lin, Z, et al.: Clustering Analysis for Silent Telecom Customers Based on K-means++, 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). IEEE, 1: 1023-1027 (2020).
10. Huang, L., Pena, B., Liu, Y., et al.: Machine learning in sustainable ship design and operation: A review. Ocean Engineering, 266: 112907 (2022).
11. IGU.: 2023 World LNG Report. https://www.igu.org/resources/lng2023-world-lng-report/ (2023)