Energy Conversion and Management: X 25 (2025) 100855

Contents lists available at ScienceDirect

Energy Conversion and Management: X

journal homepage: www.sciencedirect.com/journal/energy-conversion-and-management-x

Optimizing renewable energy site selection in rural Australia: Clustering algorithms and energy potential analysis

Iman Rahimi [a], Mufei Li [a], James Choon [a], Dane Pamuspusan [a], Yinpeng Huang [a], Bingzhen He [a], Alan Cai [a], Mohammad Reza Nikoo [b], Amir H. Gandomi [a,c,*]

[a] Faculty of Engineering and Information Technology, University of Technology Sydney, Australia
[b] Department of Civil and Architectural Engineering, Sultan Qaboos University, Muscat, Oman
[c] University Research and Innovation Center (EKIK), Óbuda University, 1034 Budapest, Hungary

ARTICLE INFO

Keywords:
Hybrid renewable energy
Solar energy potential
Wind energy potential
Optimization
Genetic algorithm
Clustering
Australia

ABSTRACT

Renewable energy development is a critical issue in Australia, and identifying suitable regions for constructing renewable energy plants is an essential step towards achieving sustainable energy goals. This work presents insights and techniques aimed at identifying optimal locations for renewable energy stations in rural areas across Australia as a whole. Following the above-mentioned idea, the study uses clustering algorithms to explore the optimization of renewable energy site selection. The research focuses on applying these algorithms to analyze spatial data and identify optimal geographic clusters for potential development based on technical parameters like solar irradiance and wind speed. Various clustering algorithms were employed in line with our methodology, namely K-Means, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Hierarchical clustering, and K-Medoids. Each algorithm generated clusters, facilitating the identification of appropriate regions based on a range of data attributes. A genetic algorithm was integrated into an iterative process to identify the most appropriate clustering method. Additionally, The HOMER Pro software was used to process the generated cluster centers and estimate the solar and wind energy potential for each location. The analysis revealed that solar panels consistently outperform wind turbines in energy generation across various clusters and algorithms. While the genetic K-Means algorithm performed best based on clustering evaluation metrics, the genetic K-Medoids algorithm produced the highest energy output. However, the latter incurred the highest financial costs, highlighting a trade-off between energy production and economic feasibility. This study provides valuable insights into the application of clustering techniques for renewable energy site selection and identifies challenges and limitations that require further investigation.

1. Introduction

As the global economy progresses, developed countries have placed substantial importance on fulfilling their renewable energy obligations. Major renewable energy sources include biomass, hydropower, geothermal, wind, and solar energy. In contrast, fossil fuels have limited availability and are linked to detrimental outcomes that contribute to unfavorable alterations, whereas the adoption of renewable energy contributes to decelerating the pace of global warming. Renewable energy sources are vital and can be viewed as feasible replacements, primarily due to their beneficial environmental impacts. As a result, they have the capacity to replace fossil fuel sources [1]. Nowadays, many countries have adopted renewable energy sources, and it is predicted that the expansion of these sustainable energy industries will continue to thrive and progress in the future [2–4]. Solar and wind energies, the crucial renewable energy sources, offer substantial long-term advantages that enhance sustainability and diminish pollution [5–8].

It has been proven that solar energy could be considered as a reliable source for generating hot water through solar photovoltaic (PV) systems, which have demonstrated their efficiency and effectiveness worldwide in recent years [9,10]. Additionally, solar thermal systems suggests the capability to harness thermal energy from the sun and are commonly employed in commercial applications [11]. It is critical to emphasize that solar panels are dependent on weather conditions, which means the energy output of photovoltaic systems depends on sunlight. Therefore, the efficiency of harnessing solar energy can be remarkably diminished

* Corresponding author.
E-mail address: Gandomi@uts.edu.au (A.H. Gandomi).

https://doi.org/10.1016/j.ecmx.2024.100855
Received 15 October 2024; Received in revised form 6 December 2024; Accepted 24 December 2024
Available online 31 December 2024

during rain or cloud cover [12]. Furthermore, the benefits of wind energy are significant, given its lack of greenhouse gas (GHG) emissions and associated global warming influences. It has the potential to generate electricity without contributing to carbon dioxide emissions [13]. While many research efforts have investigated possible locations for renewable energy ventures across Australia (e.g., [14–18]), this particular study stands out for its exclusive concentration on the distinct rural scenery of the nation. Our methodology introduces a fresh perspective on integrating renewable energy, paving the way for a transformative journey toward achieving sustainable energy self-sufficiency within these rural communities.

This study's primary contribution is identifying suitable zones for constructing energy plants in rural parts of Australia. To achieve this goal, various clustering algorithms were employed, and a genetic algorithm was used to optimize these algorithm's parameters. Through the integration of a genetic algorithm (GA) into clustering algorithms, efficient exploration of complex parameter spaces is enabled, ensuring the identification of optimal clustering configurations across multiple algorithms. This enhances the robustness and adaptability of the approach to diverse data characteristics. Moreover, several evaluation metrics were used to compare the performance of these algorithms. Subsequently, the clustering outcomes were employed to estimate different locations' solar and wind energy potential.

While numerous studies have explored renewable energy systems and site selection using clustering algorithms, there is a lack of comprehensive research that integrates multiple methods including clustering algorithms, optimization techniques, and energy simulation tools to evaluate site selection. Numerous previous studies, like [2], concentrate on particular geographical areas or a small number of approaches, frequently ignoring the relative effectiveness of clustering algorithms across different datasets and assessment criteria. Furthermore, from the literature, the majority of research does not discuss clustering parameter optimization or how it affects evaluations of renewable energy potential. The need for a more comprehensive, methodical approach to renewable energy site selection that takes into account a variety of approaches, real-world applications, and comparative performance evaluation is highlighted by this gap in the literature.

This paper addresses the above-mentioned gaps by providing a new framework that combines four clustering methods (K-Means, DBSCAN, Hierarchical, and K-Medoids) with a GA to optimize parameter selection and evaluate clustering performance. The study is further strengthened by the incorporation of HOMER Pro software[1] for energy potential simulation, which allows a direct comparison of solar and wind energy outputs across several methodologies. By using this thorough methodology in rural Australia, this study not only finds the best locations for hybrid renewable energy systems but also offers insightful information about the energy output and cost-effectiveness of various clustering strategies, which could be interesting for decision makers. These results provide a scalable methodology for maximizing renewable energy projects across many geographies, which is important for policymakers, researchers, and industry practitioners.

The remaining sections of this paper are organized as follows. Section 2 provides an overview of the literature review. In Section 3, we outline the motivation behind this study. Section 4 shows the research methodology employed. Sections 5 to 7 detail the process of data exploration, demonstrate the results and delve into discussions, respectively. Finally, Section 8 offers the conclusion of the study.

2. Literature review

Various domains have been identified for implementing renewable energy techniques, encompassing solar power, wind, rainfall, tidal energy, wave power, and geothermal warmth [3,10,19]. In the present day, a multitude of nations globally utilize renewable energy, with the anticipation that the expanding markets for such energy sources will persist and enhance robustly in the forthcoming year [4,20]. Australia is known as a developed country, which possesses the high capacity for utilizing renewable energy sources; in addition, solar power is acknowledged as a dependable renewable energy source in Australia [14,21]. Systems such as solar photovoltaic (PV) can efficiently provide hot water when using sun energy. These systems have been performing well, and in recent years, their efficacy has been recognized on a global scale [9,10,22]. Moreover, solar thermal heating and cooling systems, which capture thermal energy from the sun during daylight hours, are utilized in various commercial applications [12,23]. Regarding solar energy, however, the efficiency of solar panels is highly dependent on circumstances; which means that the availability of sunshine is a prerequisite for producing energy through photovoltaics.

Furthermore, wind energy has the least negative environmental effect when compared to other energy sources like fossil fuels [24]. While Australia still uses fossil sources, renewable sources such as solar and wind power significantly are utilized in energy and electricity generation. This is mainly because these forms of energy have numerous environmental benefits compared to their fossil fuel counterparts [25,26]. On the other side, using wind energy comes with several drawbacks; for instance, obtaining precise weather data such as wind speed and load specifications for a specific location is critical, presenting a challenge in using wind energy. Also, appropriate weather data is a prerequisite for assessing the efficiency of an existing system [27].

Moreover, due to their weather dependency, a significant disadvantage of relying solely on renewable energy sources is their inability to provide a constant energy supply. Therefore, combining these energy sources is suggested to boost the overall energy yield. Consequently, it is important to employ a suitable optimization system to determine the optimal mix of solar panels and wind turbines. An additional advantage of such a hybrid system is that combining solar and wind energy can minimize the need for battery banks and diesel. As a result, the hybrid renewable solution is ideally suited to reduce energy demands. Hybrid renewable energy systems (HRESs) commonly combine various types of renewable energy sources to improve overall system efficiency and ensure a reliable and consistent energy supply [28,29]. Identifying the optimal sites for deploying a distributed HRES poses a significant challenge, and data mining-based strategies are renowned for effectively solving this issue. Several studies employ data mining methods in this field, for instance, approaches that utilize a geographical information system (GIS) for spatial data mining [30], a data mining-based optimal demand response program [31], and other approaches [32–34].

Regarding strategizing for hybrid renewable installations, a rich and deep research in the area can be found in [35]. For example, in a study by Kazak et al., [36], a decision support system was designed to aid decision-makers in identifying suitable locations for both single-source and HRES installations to fulfill energy production needs. In a separate research, Kanata et al. [37] proposed an ideal setup and conducted a techno-economic evaluation of an HRES located on Sebesi Island within the South Lampung Regency of Indonesia. The task of identifying optimal locations for the deployment of a decentralized HRES presents a complex challenge, and the utilization of data mining-driven approaches is recognized as a potent strategy for addressing this issue [38–40]. Additional studies encompass various approaches, including a spatial data mining method based on geographical information systems (GIS) proposed by Kaundinya et al. [30], a data mining-based optimal demand response program by Babaei et al. [31], heuristic algorithms as presented by [41], and the integration of metaheuristics and clustering algorithms by [42,43]. Other noteworthy methods involve the work of [32,34].

Regarding the strategic development of hybrid renewable installations, there is an wealth of literature available on this topic [35,44,45]. In the study conducted by Kazak et al. [36], the authors introduced a decision support system designed to aid decision-makers in

[1] A free version liscence.

identifying suitable sites for both individual renewable energy sources and hybrid systems. This system aims to fulfill energy production requirements. In a different research effort, Kanata et al. [37] proposed an optimal configuration and carried out a techno-economic evaluation of an HRES implemented on Sebesi Island, located within the South Lampung Regency of Indonesia. This study introduced a holistic hybrid system design, taking into account technological, economic, and environmental considerations.

Marocco et al. [46] deliberated on the optimal configuration of off-grid HRESs, considering both the levelized energy cost (LCOE) and CO2 emissions as simultaneous factors. Various elements, including photovoltaic panels, wind turbines, batteries, hydrogen, and diesel generators, were explored to create diverse HRES setups, leading to Pareto fronts that showcase the trade-offs between costs and emissions for different configurations. Di Grazia and Tina [47] introduced a methodology that combines GIS and Multi-Criteria Decision Analysis (MCDA) for identifying ideal Fixed Photovoltaic (FPV) sites. In the aforementioned study, the authors illustrated this approach by selecting the San Giovanni Dam in Sicily from seven watersheds. This example highlighted the advantages of the proposed method, particularly for regions with high-temperature conditions. Wei et al. [48] conducted a comprehensive review that delves into the trends, models, and challenges within the realm of hybrid renewable energy research, encompassing the years 2000 to 2022. Their review involved synthesizing various energy systems, making comparisons between methodologies, tackling issues of uncertainty, and offering insights into potential directions for future advancements.

Addressing energy issues in rural and isolated locations has centered on optimizing renewable energy systems. Holloway et al. [2] recently utilized data mining, specifically the K-Means and K-Medoids clustering algorithms, to pinpoint optimal locations for deploying HRESs in rural Western Australia. While K-Medoids showed higher solar and wind energy potential in some locations, K-Means performed better overall, considering data clustering and energy requirements within clusters. The study mentioned above solely focused on one Australian state as a case study and applied two clustering algorithms. Kumar and Channi [49] delved into the feasibility of an HRES that combined PV and biomass energy sources to fulfill the energy requirements of a rural village in India.

Alavi et al. [50] showed the economic and environmental advantages of hybrid wind and hydrogen systems, by using using the ELECTRE approach, reducing CO2 emissions by 97 % when compared to conventional diesel-based systems. Similar to this study [51], used the TOPSIS technique to evaluate solar water pumping systems, ranking locations according to factors like regional demand, solar irradiation, and proximity to infrastructure. The significance of site-specific optimization strategies for guaranteeing energy sustainability and cost-effectiveness is highlighted by these studies.

The integration of hybrid energy systems into a variety of applications, including water desalination and biomass-based power plants, has demonstrated considerable capacity. Multi-Criteria Decision-Making (MCDM) techniques were used by [52] to determine the best locations for biomass plants, with a focus on regional resource usage for greatest impact. The environmental benefits of combining solar, wind, and biomass technologies were also investigated by Sadeghi et al. [53], who showed notable emission reductions. Furthermore, [54] improved the performance of hybrid systems that power reverse osmosis facilities by combining Portfolio Theory and Particle Swarm Optimization (PSO), which decreased costs and increased system reliability.

Seyed Alavi et al. [55] employed MCDM methods (TOPSIS, ELECTRE, SAW) to optimize wind farm site selection in Iran, emphasizing geographical and infrastructural factors like distance to power lines and annual wind speed. [53] optimized a hybrid solar-wind-biomass-battery energy system for rural electrification in Iran using HOMER software, achieving significant CO2 emission reductions and economic benefits. [56] proposed enhancements to particle swarm optimization (PSO) algorithms for sizing hybrid photovoltaic-diesel-battery systems in remote areas, achieving cost-effective and reliable energy supply solutions. [57] explored the optimal operation of grid-connected fuel cell-based combined heat and power systems, demonstrating the economic and environmental viability of PSO-optimized energy systems for residential use.

One creative way to democratize and decentralize energy systems is through Renewable Energy Communities (RECs). [58] presented a solution for REC planning and operation optimization according to GA approach that decreased solar energy surplus, maximized self-consumption, and shortened payback periods. Similar to the above-mentioned study, [59] created an investment optimization model to aid in REC decision-making, emphasizing operational electricity sharing and renewable energy production. Both studies presented RECs application in local energy production, usage, and financial gains.

In off-grid applications where cost and dependability are crucial, the importance of sophisticated optimization techniques in hybrid RES is clear. The performance of hybrid pumped hydro and battery storage systems was investigated by [60], who showed that they could reduce energy curtailment while maintaining a 100 % power supply. These systems were able to reach cost-effective configurations while maintaining sustainability by utilizing sophisticated algorithms such as PSO. In conclusion, these results highlight the significance of site-specific strategies, hybrid arrangements, and computational techniques in accelerating the shift to a decentralized and sustainable energy future.

While several studies have employed clustering algorithms for renewable energy site selection, including those focusing on specific regions this study offers a more comprehensive framework that combines a genetic algorithm for parameter optimization with several clustering techniques (K-Means, DBSCAN, Hierarchical, and K-Medoids). The current study provides a comparative examination of clustering approaches utilizing a comprehensive dataset covering all rural regions of Australia, in contrast to previous research that frequently concentrate on a single algorithm or small datasets. Additionally, this study is one of the first to assess the solar and wind energy potential for cluster centers that have been discovered using HOMER Pro software, allowing for a thorough comparison of energy outputs and financial expenses across the country.

Additionally, this work addresses a critical gap in the literature by optimizing clustering algorithm parameters to improve site selection accuracy and efficiency. By integrating technical attributes, such as solar irradiance and wind speed, and conducting a multi-dimensional evaluation of energy production and cost-effectiveness, this research provides a scalable methodology applicable to diverse geographies. These contributions not only advance renewable energy site selection techniques but also offer actionable insights for policymakers and stakeholders aiming to enhance energy sustainability in rural Australia.

3. Motivation

Australia's energy dilemma has been worsened in recent years by the frequency of catastrophic weather, including two more populous states, New South Wales and Victoria—projected to experience insufficient power supply. In March 2022, significant flooding in New South Wales and Queensland caused two coal mines to reduce production. Another cause of electricity outages is the cold snap. Since the cold snap hit in June, Australia has quickly transitioned into winter, the temperature has exceeded records, and there is a considerable rise in demand for electric heating. On Australia's east coast, a fifth of the coal-fired power facilities is either being repaired, shut down, or not in operation due to breakdowns, creating a 4,000 megawatt gap. Energy deprivation in rural areas of Australia is only going to be worse compared to cities. With increasing energy demand due to the growing population and reliance on fossil fuels, the use of fossil fuels increases the risks of climate change.

Rural regions in Australia are identified into three groups: large rural centers (with a population ranging from 25,000 to 99,999), small rural centers (with a population between 10,000 and 24,999), and other rural areas (with a population below 10,000). As a result, all rural areas in the country share the common feature of having a population of less than 100,000 individuals. About 28 % of Australia's total population resides in these rural towns, amounting to a collective rural population of 7,210,516 people. This computation is derived from Australia's overall population, which was reported as 25,750,200 as of September 2021 (Statistics 2020).

While rural towns in Australia possess sufficient energy connectivity and transmission infrastructure, the electricity supplied to these areas predominantly comes from non-renewable sources, particularly fossil fuels. Fossil fuels contribute to the generation of approximately 76 % of the provided electricity, with coal accounting for 54 %, gas for 20 %, and oil for 2 %.[2] Between 2019 and 2020, a mere 7 % of Australia's electricity generation came from renewable sources, underscoring the need for progress in this sector. In light of this, the primary aim of this study is to identify the best locations for hybrid renewable energy installations utilizing both solar photovoltaic and wind resources across rural Australia. The overarching objective is to maximize the utilization of renewable energy potential. Regarding the significant gap between the energy generated by solar and wind installations across Australia and the country's energy consumption, urbanized areas have been intentionally excluded from the scope of this research.

Although Australia's energy supply does include a portion derived from fossil fuels, the incorporation of clean energy sources such as solar and wind power carries significant significance in the country's energy and electricity production (Figs. 1 and 2). Previous research endeavors have placed considerable emphasis on evaluating appropriate locations for deploying renewable energy systems (RESs) in the rural regions of Australia. This study seeks to make a substantial contribution by addressing a notable research gap concerning optimizing HRESs. The primary objective of this study is to pinpoint prospective locations for implementing decentralized hybrid renewable energy generation systems in rural regions throughout Australia. While prior studies have delved into potential sites for renewable energy systems in various rural areas of Australia [2,61–66], the present research aims to offer a more thorough and detailed examination of this subject.

To the best of our knowledge, this research marks the inaugural dedicated endeavor concentrated solely on rural regions spanning the entire nation of Australia. Our research aims to underscore not only the technical viability and energy generation potential of HRESs but also their potential to bring about favorable effects on local communities, economies, and the environment. By scrutinizing locations characterized by abundant solar radiation and strong wind speeds to optimize energy generation, our intent is twofold: to contribute to the seamless integration of renewable energy and to illuminate the considerable benefits that these systems bring. These advantages encompass diminished electricity expenses, reduced upkeep expenditures, and curbing greenhouse gas emissions. Our concentration on rural Australia is motivated by the aspiration to bridge this gap in existing literature. Our objective is to furnish a thorough examination that encompasses not only the technical intricacies but also the socio-economic dimensions of integrating renewable energy.

Rural regions generally exhibit lower energy requirements compared to suburban areas, creating a viable scenario for meeting the energy needs of rural communities solely through renewable sources. Consistent with this goal, our objective was to pinpoint the most appropriate locations—those characterized by steady solar irradiation and intense wind speeds—to install hybrid renewable energy facilities that integrate both solar photovoltaic and wind energy sources. In particular, we focused on select rural zones within Australia. Leveraging the HOMER software[3] (LLC, n.d.) and employing various clustering methodologies, we conducted an analysis to pinpoint areas with the potential to sustain HRESs capable of providing ample power to the surrounding vicinity. The present study encompasses the following inquiries:

- Which regions in rural Australia encounter substantial solar irradiation?
- Which areas in rural Australia have elevated wind speeds and frequent wind occurrences?
- How does the application of diverse clustering methods influence the arrangement of clusters?
- What is the expected energy production (measured in kWh/yr) from HRESs in each cluster?
- Which locations within Australia are best suited for the installation of HRESs?

To summarize, this study aims to fill a knowledge void within rural Australia by offering valuable perspectives on the feasibility of incorporating renewable energy sources. This endeavor contributes to sustainable progress and self-reliance in terms of energy for rural communities in the area.

4. Research methodology

To achieve the research objectives, the method we will undertake to answer the research problem will include: analysis and filtration of the Australian Towns Dataset, Transform data into format readable by clustering algorithms, implementation of clustering algorithms, optimization algorithm using GA, comparing the algorithms using metrics, input centroids locations into HOMER PRO software for processing, evaluate and analyze results. In accordance with this methodology, we will collect data from and analyze the data presented from the Australian towns dataset, which contains necessary data about Australian towns. Given the unique traits of our data and research objectives, integrating these techniques was deemed the most effective approach. We incorporated a GA in an iterative process to identify the most suitable clustering method. This approach, we believe, was the most fitting given the high-dimensional and potentially noisy nature of our data.

As the desired solution does not require a modelling technique, clustering algorithms will be implemented for the solution presented. The chosen clustering algorithms include: K-Means, DBSCAN, Hierarchical clustering, and K-Medoids. Each algorithm will construct a total of 10 clusters ($K = 10$), as a uniform test for each algorithm. As a further form of analysis, each algorithm will be optimized as a GA. The purpose of implementing GA in the clustering algorithms is to further optimize the algorithms themselves to produce a more suitable result. GAs optimize a solution by searching for the highest score-producing solution based on its fitness function, operating on a number of generations. Within each generation, the algorithm obtains the best solution and alters its parameters randomly to find its next best solution. Finally, each algorithm will be evaluated using three unsupervised machine learning evaluation metrics, namely: Silhouette [67], Davies Bouldin (DB) [68], and Calinski Harabasz's scores [69] (Fig. 3).

The method we will undertake to answer the research problem is as follows: Analyzing and filtering the Australian Towns Dataset, Transforming data into a format readable by clustering algorithms, implementing clustering algorithms, optimizing algorithm output using GA, comparing the algorithms using metrics, inputting centroid locations into HOMER PRO software for processing, evaluating and analyzing results. Each genetic algorithm clustering technique's resulting cluster centres output will form the input into the HOMER Pro software. The HOMER Pro software will generate the cluster centres' potential solar

[2] Australian Government: Department of Industry, Science, Energy and Resources. (2020). States and territories. https://www.energy.gov.au/data/states-and-territories.

[3] Free trial version (3.15.3_x64).

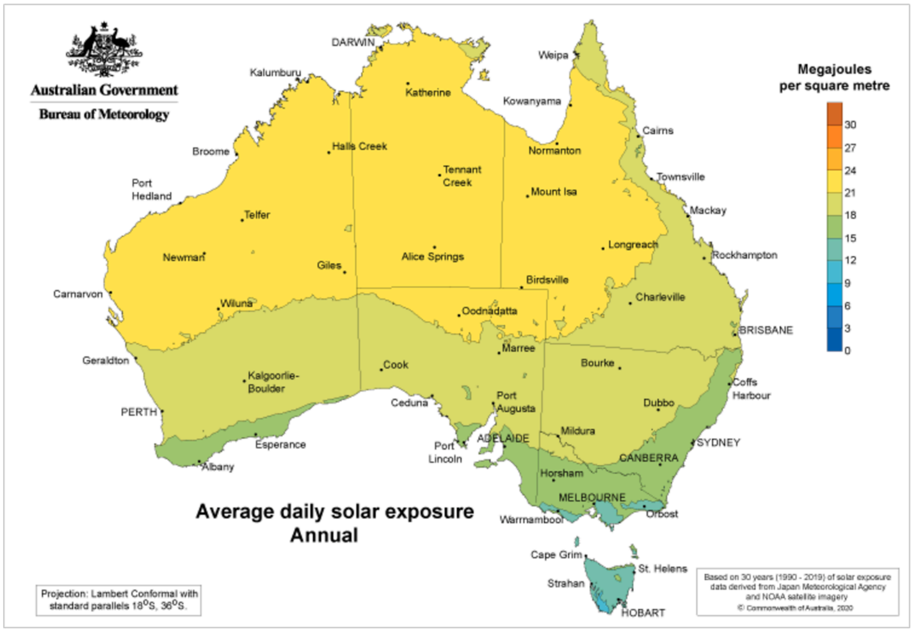

Fig. 1. Average daily solar exposure (http://www.bom.gov.au/climate/maps/averages/solar-exposure).

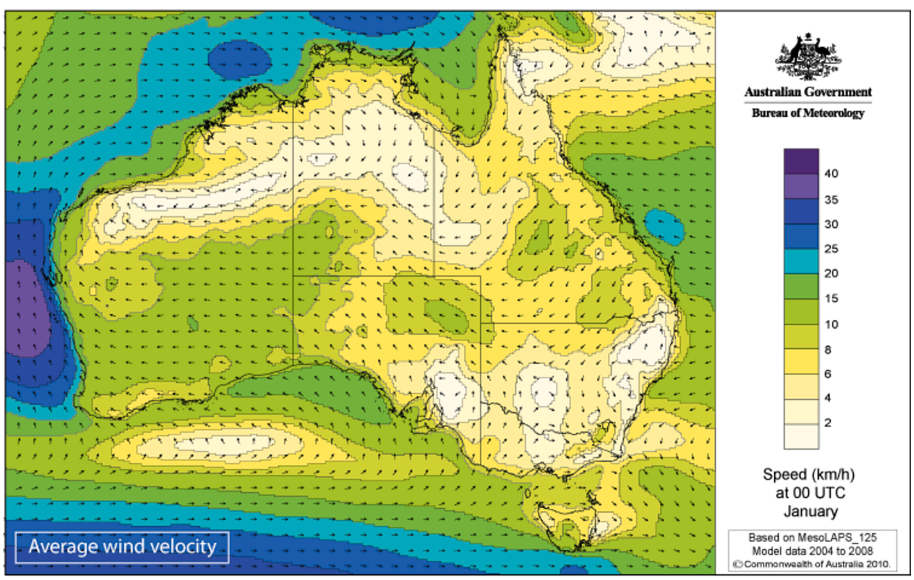

Fig. 2. Average wind velocity (http://www.bom.gov.au/climate/maps/averages/wind-velocity).

and wind energy [2]. The resulting energy output will be compared for each algorithm according to its efficacy as a solution to the problem.

4.1. Clustering algorithms

The employed clustering methods in this study are defined as follows:

- K-means clustering: In the implementation, the scikit-learn library's implementation of K-means is utilized and employed methods such as the elbow method or silhouette analysis to determine the optimal value of 'k'. The appeal of K-means lies in its simplicity and efficiency. It neatly partitions datasets into 'k' clusters. It performs best when the clusters are spherical and evenly sized, and when the dataset is devoid of outliers. Therefore, careful consideration must be given to the selection of 'k' and the initial centroids.
- DBSCAN: This technique excels in handling spatial clusters of various shapes and sizes and effectively managing noise and outliers. Given the heterogeneous density within our dataset, DBSCAN was chosen for its adaptability in forming clusters. The algorithm's parameters, such as epsilon (eps) and the minimum number of samples (min_samples), are tuned based on domain knowledge and visual inspection of the data.
- Hierarchical clustering: This algorithm offers a distinct advantage as it does not mandate the specification of the number of clusters upfront. It begins by treating each data point as an individual cluster, and then sequentially merges these based on similarity, resulting in a dendrogram of clusters. However, a caveat of hierarchical clustering is its irreversibility – once two clusters are combined, the decision cannot be undone, which might lead to less-than-optimal results. The agglomerative hierarchical clustering algorithm from the sci-kit-learn library is employed, and dendrograms are used to visualize

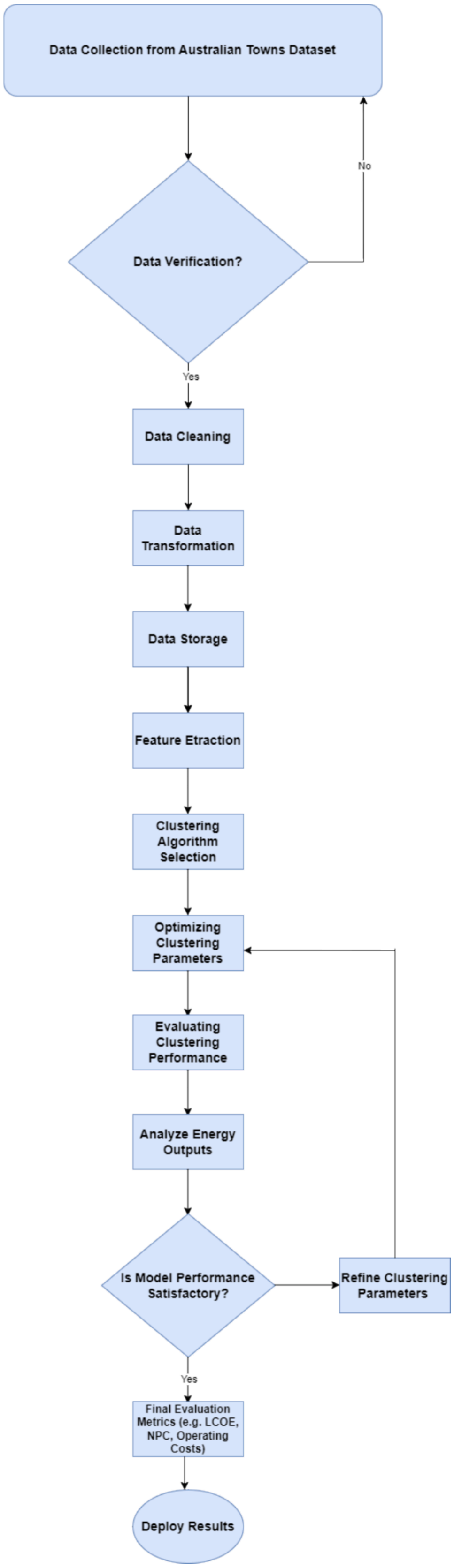

Fig. 3. Flow Chart of the Research Methodology.

the hierarchical clustering process and determine the appropriate number of clusters.

- K-medoids clustering: K-medoids were chosen for their robustness in scenarios where dataset stability is crucial. This technique suggests as a robust alternative to K-means. It shines when the dataset contains outliers or non-normal distributions, thanks to its use of actual data points rather than means as cluster centers. Despite being computationally more intensive, its robustness makes it an excellent choice when dataset sturdiness is paramount.

The clustering process in this study was conducted using Python programming, with clustering algorithms implemented through libraries such as Scikit-learn and Scikit-learn-extra. Python programming was used to implement the clustering procedure in this study, and Scikit-learn and Scikit-learn-extra libraries were used to conduct the clustering methods. The K-Means, DBSCAN, Hierarchical, and K-Medoids algorithms were executed more easily on spatial data from the Australian Towns Dataset thanks to these technologies. Latitude and longitude coordinates were the main spatial inputs for clustering, which were supplemented by other characteristics like population and land area to fine-tune cluster allocations. Raster data inputs, including solar radiation and wind speed obtained from NASA's Prediction of Worldwide Energy Resources, were preprocessed and aligned with the spatial point data format. Python tools such as Matplotlib and Folium were used to visualize the clusters produced by these algorithms. The centroids of the resultant clusters were then fed into the HOMER Pro software to assess the potential for renewable energy and improve site selection.

Table 1 provides a comprehensive overview of the parameters utilized in both the GA and the clustering algorithms. It outlines key aspects such as the number of clusters (k), the number of generations, the selection type, crossover type, crossover probability, mutation type, mutation replacement, mutation rate, and other relevant parameters. This concise summary enhances the transparency of the methodology, aiding readers in understanding the approach and facilitating an assessment of its robustness.

4.2. Evaluation metrics

In the realm of clustering, gauging the efficacy of a particular method is a complex task. Hence, we adopted a quartet of distinct evaluation metrics to compare and contrast the effectiveness of each hybrid algorithmic approach. These metrics provide invaluable insights into the various aspects of the clustering output. They help us to assess the compactness of our clusters, the degree of separation between them, and the appropriateness of the number of clusters formed. Here's a snapshot of the metrics employed:

4.2.1. Silhouette score

Cluster evaluations rely on the silhouette score, which assesses the degree of similarity between an object and its own cluster relative to other clusters. The silhouette score falls within a spectrum from -1 to 1, where higher values signify superior clustering outcomes.

4.2.2. Davies Bouldin score

Much like the Silhouette Score, the DB Score serves as a metric for evaluating clustering quality. It assesses clustering quality by considering both cluster separation and cluster compactness. Lower values of the DB Score indicate superior clustering results, indicating improved cluster separation and compactness.

4.2.3. Calinski Harabasz score

The Calinski-Harabasz (CH) score quantifies the accuracy of clustering outcomes. Elevated CH scores denote superior clustering results, and they are computed based on the ratio of within-cluster variance to between-cluster variance.

Table 1
Algorithm Parameters Summary.

Algorithms Parameters	GA − Kmeans Clustering	GA − DBSCAN Clustering	GA − Hierarchical Clustering	GA − K −Medoids Clustering
Clustering Algorithm Library in Python	sklearn.cluster.KMeans	sklearn.cluster.DBSCAN	sklearn.cluster.AgglomerativeClustering	sklearn_extra.cluster.KMedoids
Genetic Algorithm Library	pygad	pygad	Custom genetic algorithm implementation	pygad
Number of clusters (k)	10	Not applicable (DBSCAN is a density-based clustering algorithm)	10	10
Number of generations	1000	100	10	100
Number of parents mating	5	5	50 (implicit, calculated as half of the population size)	10
Solutions per population	50	50	100	50
Number of genes	k * X.shape[1]	2 (eps, min_samples)	Not applicable	Not applicable(KMedoids doesn't use genes; it directly selects cluster centers, known as Medoids, from the dataset.)
Gene type	float	float (for eps) int (for min_samples)	Not applicable	Not applicable
Gene value range	−45 to 160	eps: Randomly selected from 0.1 to 1; min_samples: Randomly selected from 2 to 10	Not applicable	Not applicable
Selection type	steady state selection	steady state selection	implicit selection based on fitness value silhouette_score	steady state selection
Crossover type	single_point	Not applicable (crossover operation not explicitly defined for GA-DBSCAN)	Not applicable	Not applicable
Crossover probability	0.25	Not applicable	Not applicable	Not applicable
Mutation type	random	random	Not applicable	None (Mutation is disabled in this type)
Mutation replacement	True	True	Not applicable	Not applicable
Mutation value range	−45 to 160	eps: Mutated by adding a random value between −0.05 and 0.05; min_samples: Mutated by adding a random integer between −1 and 1	Not applicable	Not applicable
Mutation rate	10 %	10 %	Not applicable	Not applicable

4.2.4. *Differences*

The Silhouette Score places importance on both intra-cluster similarities and inter-cluster dissimilarities. In contrast, the Davies-Bouldin score focuses on cluster separation and compactness. Meanwhile, the CH score highlights the ratio of variance within clusters to variance between clusters. Each of these methods has its own set of advantages and disadvantages, and the choice of which method to use depends on the specific problem. It is important to note that these evaluation metrics have their unique strengths and limitations, and their effectiveness hinges on the specifics of the data and the study's requirements. In the subsequent sections of this work, we will dig deeper into these metrics' underpinnings, shedding light on how they have guided our quest toward the most appropriate clustering technique for this study.

4.3. *HOMER Pro software*

HOMER is widely recognized and used in both academic and professional settings for hybrid renewable energy systems. Its extensive use across various studies in both academia and industry provides a solid foundation for comparability and validation against established benchmarks in renewable energy research [35,70]. The software offers comprehensive tools for simulating various energy sources and their interactions, which are crucial for analyzing hybrid systems. Moreover, HOMER Pro's ability to model different configurations and sensitivity scenarios allows us to explore a range of potential outcomes and identify robust strategies for rural energy deployment [71–73].

The current study suggested a combination of wind turbines and solar PV panels as its main energy sources, with battery storage for peak load control and energy dependability. Through the use of an inverter, the solar PV panels transform sunlight into direct current (DC) power, which is subsequently converted into alternating current (AC) for grid compatibility. Direct AC power generation from wind turbines is either used right away or transformed into DC for lithium-ion battery storage.

The battery storage system makes up for energy losses during times of low wind or solar irradiance, ensuring a steady supply of power. The system's operational longevity is also increased by a charge controller, which controls the battery charging process to avoid overcharging or discharging beyond safe limits.

To maximize power distribution among energy generation, storage, and load demand, the system uses an energy management controller. While preserving grid stability, this controller gives priority to the utilization of renewable energy. The excess electricity is either exported to the grid or stored in the batteries when the output of renewable energy surpasses the use. On the other hand, the energy stored in the batteries is used to meet load requirements during times of high demand or decreased generation. By ensuring a balanced and effective energy flow, this integrated strategy reduces reliance on non-renewable resources and improves the sustainability of the system.

4.4. *Numerical analysis domains*

The numerical analysis domains in this study focus on evaluating clustering algorithms and hybrid renewable energy system performance. Clustering techniques (e.g., K-Means, DBSCAN) using latitude, longitude, sun irradiance, and wind speed are used in spatial data analysis to categorize regions, which are then displayed through geographic mapping for regional suitability across the country. Moreover, Energy Production Analysis simulates solar and wind energy outputs using HOMER Pro. Average wind speeds are used to gauge wind turbine efficiency, while Global Horizontal Irradiance (GHI) is used to gauge solar panel performance. In order to assess viability, economic feasibility takes into account indicators such as operational costs, Net Present Cost (NPC), and Levelized Cost of Energy (LCOE). In order to optimize site selection, Algorithmic Performance assesses the compactness and separation of clusters using metrics such the CH Score, Davies-Bouldin Index, and Silhouette Score.

5. Data exploration

The dataset used for analysis and clustering in this work is the Australian Towns List[4]. Regarding the data source, the Australian Towns dataset from the Ready-to-use List of Australian Towns has been utilized, sourced from the website https://www.australiantownslist.com/. This dataset encompasses detailed information on all 15,323 cities, towns, villages, and suburbs in Australia, including state, postcode, latitude-longitude, local government, region, population, and more. The data is collected from the Australian Bureau of Statistics, Geoscience Australia, and the latest census releases, with updates made on April 22, 2023. The website pledges a comprehensive update of the data every quarter, subjected to regression testing and quality assurance checks before release, and offers real-time online search for data testing. Moreover, the website has been operational since 2012, serving hundreds of Australian or international businesses, allowing them to freely utilize the data for commercial purposes. We opted for this dataset because it provides the most comprehensive, accurate, timely, and commercially usable information on Australian towns, meeting the requirements of our research. Table 2 provides an overview of the data layers utilized in this study, including their spatial resolution, sources, and types. This summary clarifies the structure and attributes of the datasets, facilitating a better understanding of the methodology applied.

In order to gain access to the more extensive dataset, as it includes geographic details such as place name, urban area, state code, state, postcode, latitude, longitude, population, median income, elevation, area (km^2), local government area, region, time zone, and type. While not all the attributes are required for use in the discovery stage, the data is relevant for use in the exploration stage and are defined as follows:

5.1. Data attributes

The most relevant data attributes used for both exploration and implementation are the attributes pertaining to a location: Name, State, Latitude, Longitude, Population, Area (km^2), Region and Type.

Name: Indicates the location's official name as a text string of a maximum of 37 characters. Each name follows a strict naming convention as specified by the dataset creators.

State: Presents the full name of the Australian State, can only be a single value, formatted as a string, i.e., New South Wales or Victoria, etc.

Latitude and Longitude: The WGS83 Latitude and Longitude coordinate value for the location's center point. Takes a float value of 5 decimal place precision.

Population: Integer value of the number of people who reside in the location record as found in the 2021 census.

Area: Floating point value of the Area of a location represented in km^2. Uses 3 decimal place precision.

Region: A string value represents the ABS Level 4 Statistical Area under which the location is classified.

Type: A string value indicating the category of location can be one of the following options: "Rural locality," "Urban locality," "Major urban locality" (with a population exceeding 20,000), or "Suburb."

5.2. Preprocessing

The preprocessing process starts with filtering out unnecessary data for use in the implementation. Specifically, locations not classified as "Rural locality" under the attribute type are omitted. A further search of the data found that there were locations with 0 population or greater than 10,000; these locations were omitted since they were deemed irrelevant as those with 0 population would not require energy resources, and those with 10,000 people or greater are not considered in the constraints of this work. Another interesting finding was the inclusion of Norfolk Island, Christmas Island and Lord Howe Island. In the dataset, these locations are off the mainland of Australia and would not necessarily benefit from the energy sites developed in the solution. Moreover, they may affect the clustering algorithm, which would use its outlying longitude and latitude attributes to develop its clusters. Finally, two towns with a recorded 0 km^2 area were omitted. This process was aided by Microsoft Excel, of which each preprocessing step occurred in a new sheet (Supplementary file).

The purpose of this filtration of data is due to the constraints of the research problem as well as the practicality of the energy station installation. The constraints outline that these sites are to be in rural areas with low populations and enough space to install solar panels and wind turbine sites.

Following the exclusion of non-rural locations and towns with populations exceeding 10,000, the dataset retained a total of 9,625 towns. Among these, the majority of suitable towns, numbering 2,774, are situated in New South Wales, with Queensland (2,094) and Victoria (2,006) following closely. Additionally, Western Australia and South Australia each have 1,033 and 1,044 towns, respectively, meeting the stipulated criteria, while Tasmania accounts for 525 towns that fit the conditions (Fig. 4).

Regarding population size, a substantial portion of regions, specifically 6,705 regions, have populations ranging from 0 to 150 individuals. These regions satisfy the fundamental criteria for the installation of a power plant. Most of the towns in the dataset consist of this attribute, making the data suitable for the study implementation (Fig. 5). Other notable data points include:

- Highest Latitude: Ugar Island, Queensland (−9.51)
- Lowest Latitude: Recherche, Tasmania (−43.54)
- Highest Longitude: Broken Head, New South Wales (153.59)
- Lowest Longitude: Dirk Hartog Island, Western Australia (113.05)
- Highest Elevation: Charlotte Pass, New South Wales (1837)
- Lowest Elevation: The Percy Group, Queensland (−19)
- Largest Population: Kialla, Victoria (8,667)
- Largest Area: Telfer, Western Australia (178,407 km^2)

6. Results

In order to visualize the algorithm's output practically, we utilized a combination of Matplotlib and folium. Map libraries. Additionally, we used the geoJSON map data from GitHub for Australian states and territories; this data is used alongside the folium. Map library as a map background that can display the longitude and latitude data of each town site, additionally showing the centre point of each cluster[5].

6.1. HOMER PRO

The HOMER software is used to compare the clustering centers obtained by different algorithms. Enter a coordinate into HOMERPRO and download solar and wind resource information for that coordinate from "NASA Prediction of Worldwide Energy Resources", we will be able to predict how much renewable energy we can get from local solar panels and wind turbines.

For example, we choose (−38.86365047,146.8174897) this coordinate, and we can obtain this information below (Figs. 6-8). We will introduce more information in the Finding Section.

Below are the specified settings for the configuration.

- Wind Turbine: Generic 3 kW
- Solar Panel: Generic Flat Plate PV
- Battery: Generic 1-kWh Lithium-Ion Battery

[4] australiantownslist.com.

[5] <https://github.com/rowanhogan/australian-states>.

Table 2
Overview of Data Used.

Layer	Spatial Resolution	Source	Layer Type	Description
Town Locations	Point Data	Australian Towns Dataset	Vector (Point)	Latitude and longitude coordinates of towns in Australia.
Population Data	Administrative Level	Australian Bureau of Statistics	Attribute Table	Population data from the 2021 census for each town.
Geographic Boundaries	1:100,000	Australian Towns Dataset	Vector (Polygon)	Boundaries of Australian states and territories.
Solar Radiation	Global	NASA Prediction of Worldwide Energy Resources	Raster	Global Horizontal Irradiance (GHI) data for solar energy potential assessment.
Wind Speed	Global	NASA Prediction of Worldwide Energy Resources	Raster	Average wind speed data for renewable energy evaluation.
Land Area	Administrative Level	Australian Bureau of Statistics	Attribute Table	Area in km^2 for each town or locality.

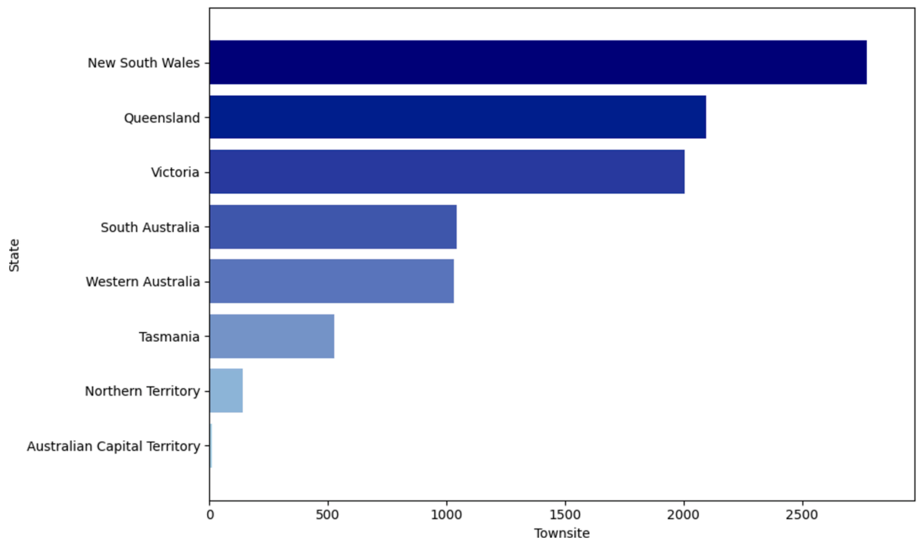

Fig. 4. The Distribution of townsites in different Australian States.

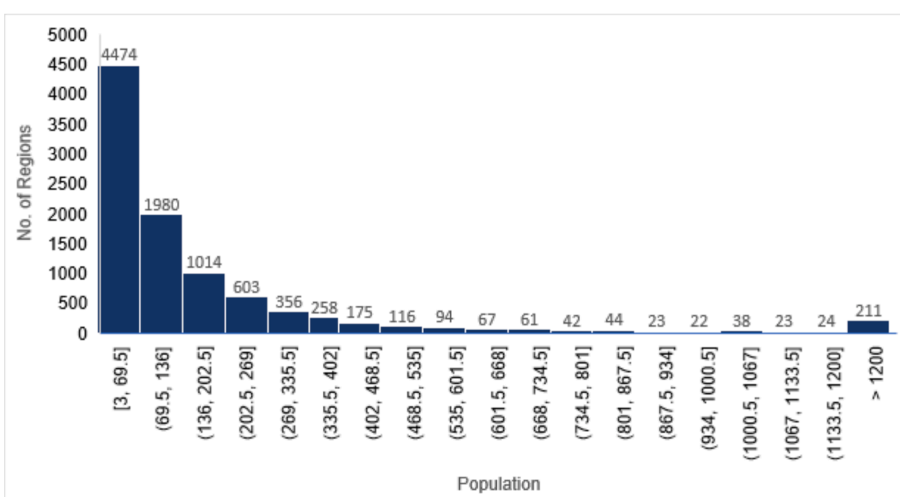

Fig. 5. The Population Distribution in Twnsites.

We intentionally chose specific parameters, such as the Generic 3 kW Wind Turbine, Generic Flat Plate PV solar panel, and Generic 1-kWh Lithium-Ion Battery, in order to strike a balance between representing commonly used technologies and conducting a feasibility assessment within the scope of our study. While there is the prospect of a design process geared towards maximizing renewable sources in identified clusters, it is crucial to recognize our research's practical constraints and objectives. Our main focus was evaluating the potential energy output and feasibility of hybrid renewable energy systems in specific rural areas of Australia. By employing standard values for the wind turbine, solar panel, and battery, we aimed to provide an initial analysis of energy potential rather than exhaustively optimizing each component.

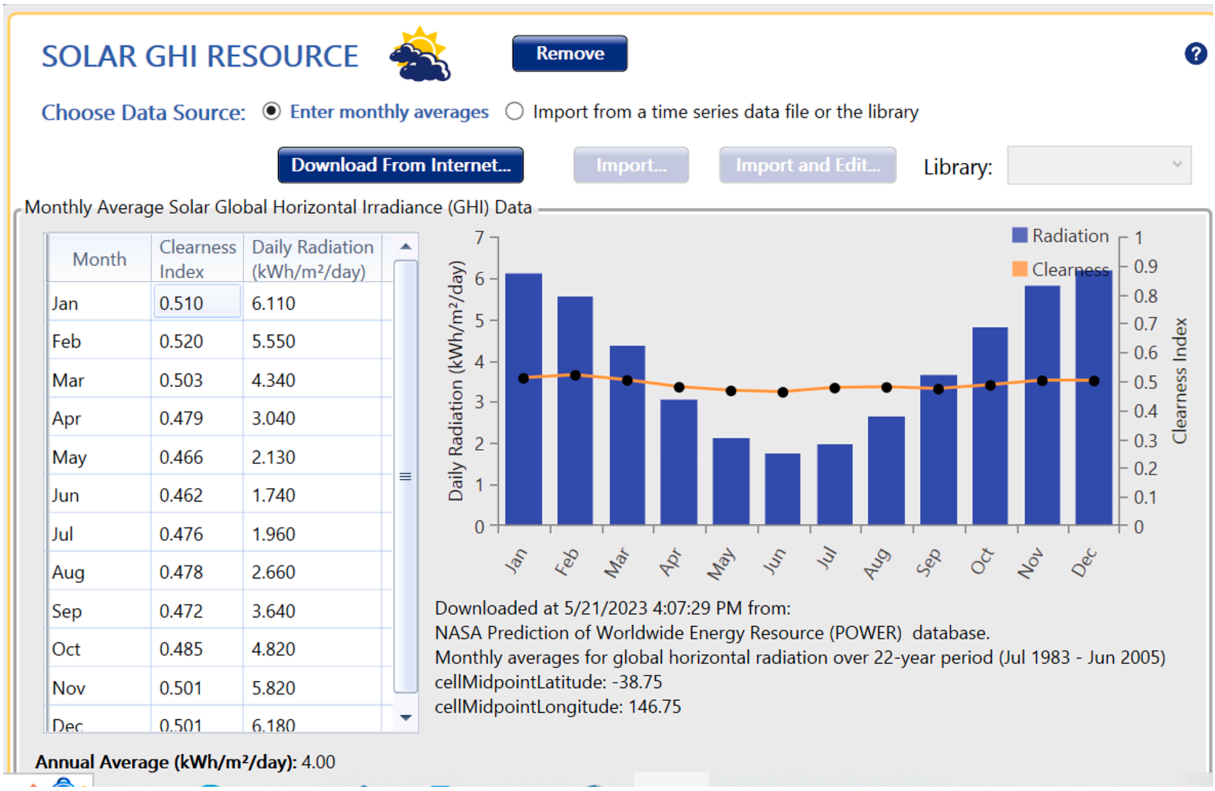

Fig. 6. Solar GHI Resource of location (−38.86365047,146.8174897).

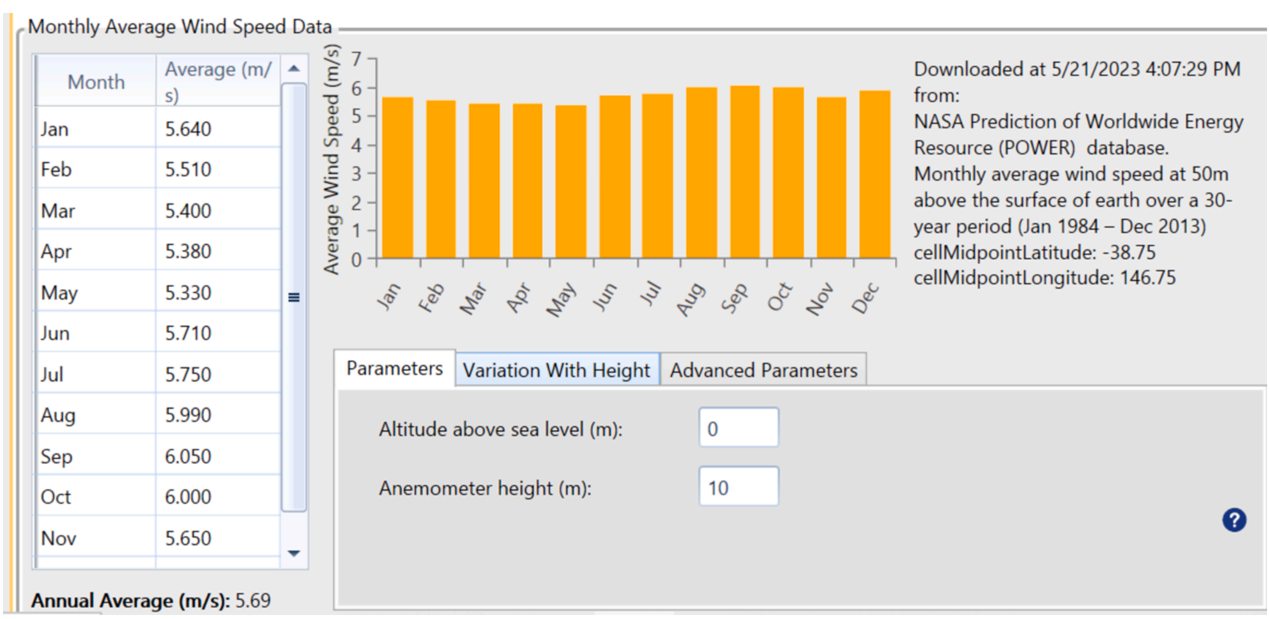

Fig. 7. Wind Resource of location (−38.86365047,146.8174897).

6.2. *Visualize the clustering results*

In alignment with our chosen approach, we intend to collect and evaluate data from the Australian Towns Dataset, which provides fundamental details about towns in Australia. Given the unique attributes of our data and our research objectives, we have opted for a strategy involving diverse methods. Specifically, a GA has been incorporated into a repetitive procedure to determine the most optimal clustering technique. We hold the view that this method is the most fitting one, considering the intricate and possibly turbulent characteristics of our data, which is characterized by a high number of dimensions.

As our desired solution does not require a specific modelling technique, we will implement clustering algorithms for the proposed solution. The chosen clustering algorithms include K-Means, Hierarchical clustering, DBSCAN, and K-Medoids. For consistent evaluation, each of these algorithms namely, K-Means, Hierarchical clustering, and K-Medoids will generate ten clusters (K = 10) while DBSCAN

Optimization Results

Left Double Click on a particular system to see its detailed Simulation Results.

Architecture					Cost				System		PV		
PV (kW)	G3	1kWh LI (#)	Converter (kW)	Dispatch	NPC ($)	LCOE ($/kWh)	Operating cost ($/yr)	CAPEX ($)	Ren Frac (%)	Total Fuel (L/yr)	CAPEX	Energy Production (kWh/yr)	Capital ($)
2.66	1	19	2.13	CC	$47,197	$0.889	$885.83	$35,745	100	0	6,655	3,462	18,000
7.88		42	4.58	CC	$59,060	$1.11	$1,151	$44,182	100	0	19,707	10,254	
	2	49	1.84	CC	$88,971	$1.68	$1,970	$63,502	100	0			36,000

Fig. 8. Optimization Results of Installing Solar Panel and Wind Turbine in this location.

automatically determines clusters based on data density. To further enhance their performance, we will optimize each algorithm using a GA from the Scikit-learn library as an additional analytical step. Genetic Algorithms aim to improve solutions by searching for the highest-scoring solution based on their fitness function over multiple generations. The algorithm identifies the best solution in each generation and introduces random parameter adjustments to find the next best one. Ultimately, we will assess the performance of each algorithm using three evaluation metrics: Silhouette, DB, and CH scores.

Our approach to addressing the research problem will involve several steps:

- Analyzing and filtering the Australian Towns Dataset.
- Transforming the data into a format compatible with clustering algorithms.
- Implementing clustering algorithms.
- Optimizing algorithm outputs using GA.
- Comparing the algorithms using metrics.
- Feeding centroid locations into HOMER PRO software for further processing.
- Evaluating and analyzing the results.

The cluster centers produced by each GA clustering technique will be used as input data for the HOMER Pro software. HOMER Pro software will compute these cluster centres' potential solar and wind energy. The resulting energy outputs will be compared among the algorithms to evaluate their efficacy as potential solutions to the problem, and these findings will be presented in Figs. 9-12. The GA is integrated with four distinct clustering algorithms: K-Means, DBSCAN, Hierarchical, and K-Medoids.

The difference in the number of clusters generated by the GA-DBSCAN method compared to other proposed clustering methods in this study arises from the inherent characteristics of DBSCAN as a density-based clustering algorithm. In DBSCAN algorithm, factors such as the neighborhood radius and the minimum number of points are used to create a cluster to determine clusters based on density, as opposed to K-Means or K-Medoids, which generate a predetermined number of clusters. Based on the spatial data properties, the GA-DBSCAN algorithm in our study maximized these parameters to produce the most relevant grouping. Only two clusters (0 and 1) were formed in the final analysis due to the input data's spatial dispersion and density. This shows that other areas were classified as noise or outliers, but only two dense regions met the clustering requirements. Because clusters with fewer points may have less potential for resource-sharing or optimization, this result has a direct impact on cost-related metrics, such as operational expenses and LCOE. It is crucial to remember that DBSCAN's adaptability in identifying organic clusters, independent of predetermined cluster numbers, offers special insights into the density-based properties

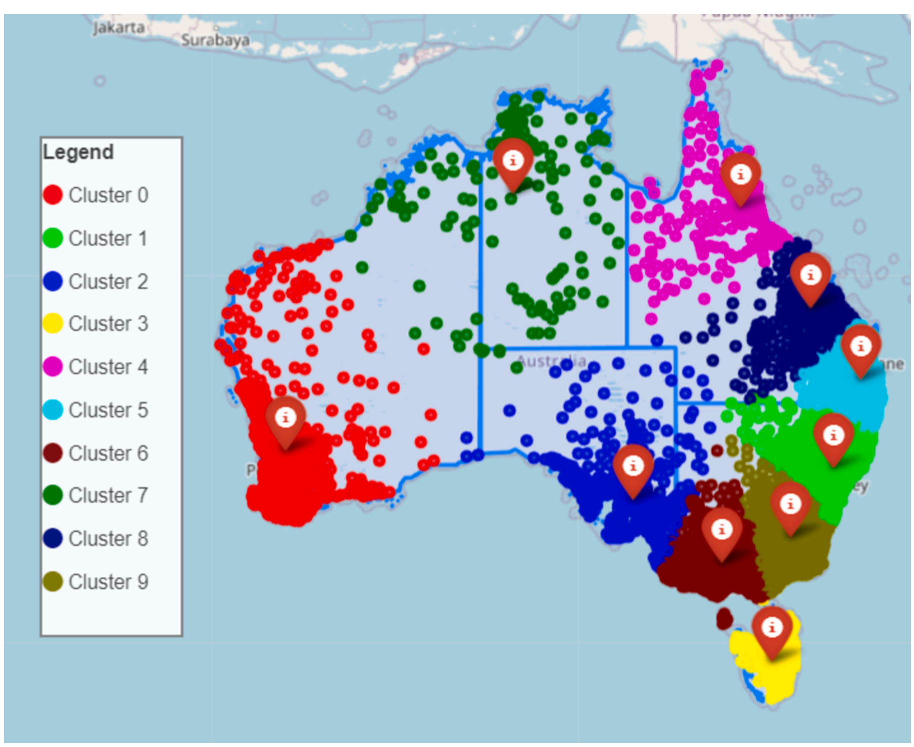

Fig. 9. Results of the GA – Kmeans method.

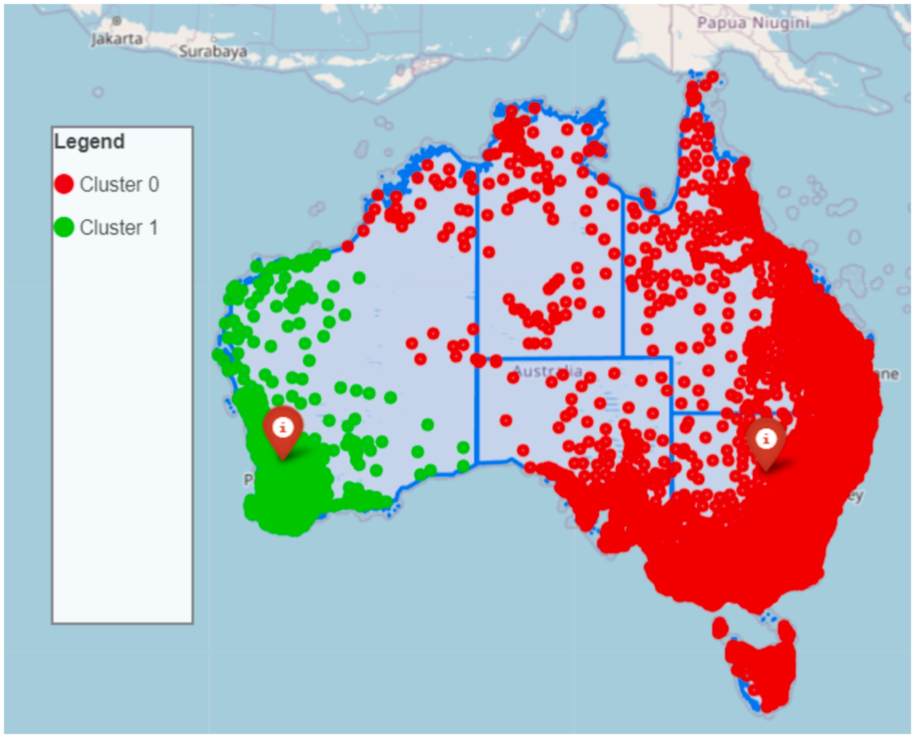

Fig. 10. Results of the GA – DBSCAN method.

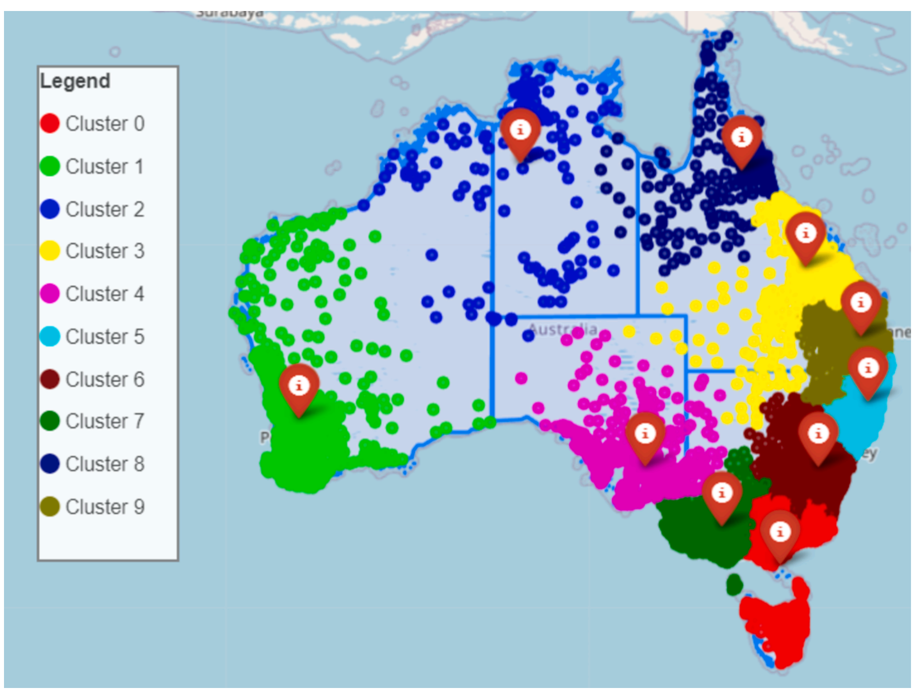

Fig. 11. Results of the GA – Hierarchical method.

of the spatial data. Details of the clusters (Latitude and Longitude) have been shown in Tables 3-6.

7. Discussion

The following subsections present our findings and initiate discussions. First, we conduct a comparative analysis of the algorithms using evaluation metrics. Subsequently, we compare energy production based on the results. Additionally, we provide an overview of costs, and finally, we summarize our findings.

7.1. Comparison study

The following subsections offer a comparative study covering evaluation metrics, energy production, and costs.

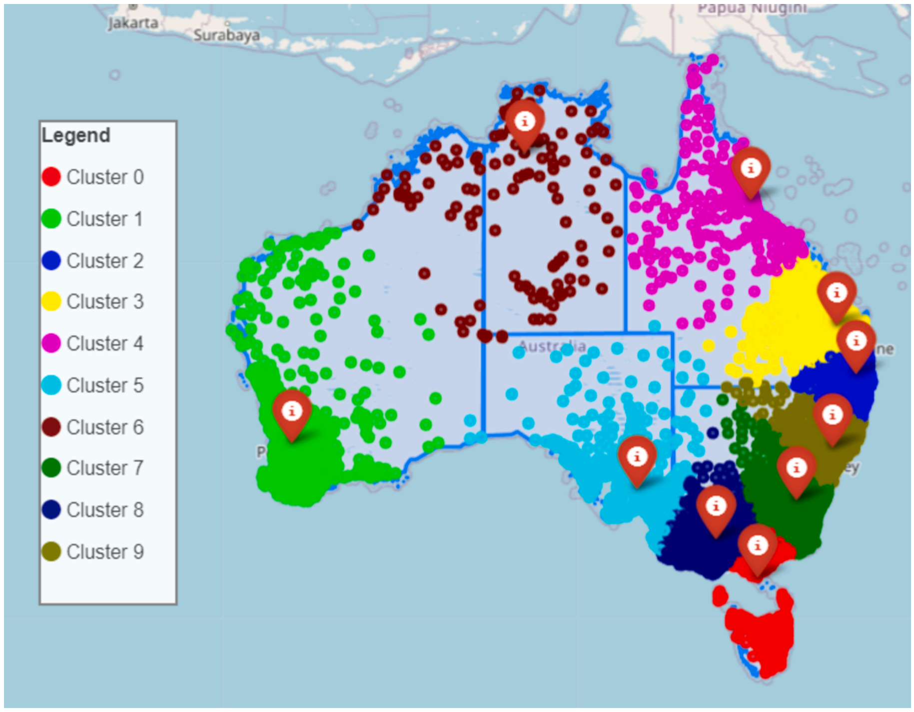

Fig. 12. Results of the GA – K – Medoids method.

Table 3
Cluster details for GA-Kmeans.

Cluster	Latitude	Longitude
Cluster 0	−32.25	116.49
Cluster 1	−31.51	151.27
Cluster 2	−33.62	133.94
Cluster 3	−38.01	145.34
Cluster 4	−18.05	145.86
Cluster 5	−27.17	152.36
Cluster 6	−34.72	139.21
Cluster 7	−16.8	135.82
Cluster 8	−24.24	149.46
Cluster 9	−34.98	149.07

Table 4
Cluster details for GA-DBSCAN.

Cluster	Latitude	Longitude
Cluster 0	−32.12	146.82
Cluster 1	−31.52	116.96

Table 5
Cluster details for GA-Hierarchical.

Cluster	Latitude	Longitude
Cluster 0	−38.86	146.82
Cluster 1	−31.51	116.99
Cluster 2	−17.15	130.73
Cluster 3	−23.17	148.35
Cluster 4	−34.02	138.42
Cluster 5	−30.62	152.21
Cluster 6	−34.02	149.11
Cluster 7	−37.03	143.19
Cluster 8	−17.59	144.42
Cluster 9	−27.08	151.79

Table 6
Cluster details for GA-KMedoids.

Cluster	Latitude	Longitude
Cluster 0	−38.95	146.36
Cluster 1	−32.02	116.81
Cluster 2	−28.23	152.56
Cluster 3	−25.46	151.31
Cluster 4	−18.12	145.91
Cluster 5	−34.42	138.7
Cluster 6	−15.3	131.57
Cluster 7	−35.02	148.84
Cluster 8	−37.03	143.68
Cluster 9	−32.28	151.11

7.1.1. Evaluation metrics comparison

The results generated from the GA implementations were altered slightly using the updated dataset and adjustments to the implementation of the evaluation metrics. The updated results are displayed in Figs. 13-14 below. Based on the initial evaluation metrics, the algorithm that performed most effectively by both DB and CH scores was GA-Kmeans.

In order to evaluate the results of each algorithm mentioned, the algorithms were compared using the Friedman and Post-Hoc Nemenyi test. The Freidman test uses a nonparametric method that evaluates the similarity of algorithms by first analyzing the significance of the difference in the performance of the algorithms to be evaluated. The output is the test statistic of the null hypothesis test, which determines if the algorithms are consistent. The Post-Hoc Nemenyi test evaluates the effectiveness of the algorithms using the p-value of the Friedman test, which is 0.978 (test statistic 0.200), which is statistically significant; this means that the post hoc test can determine which groups of results are different from each other. It can be observed from the implementation, the test returns the p-values of a pairwise comparison of the means of the data. The two groups with statistically significant different means from the matrix generated are 0 and 2 (Table 7).

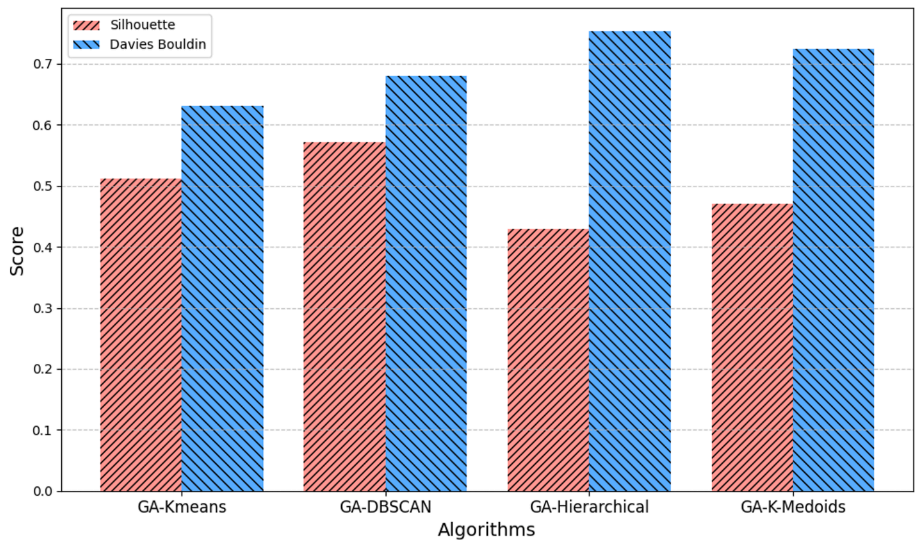

Fig. 13. Comparing Silhouette and Davies Bouldin Score Gained.

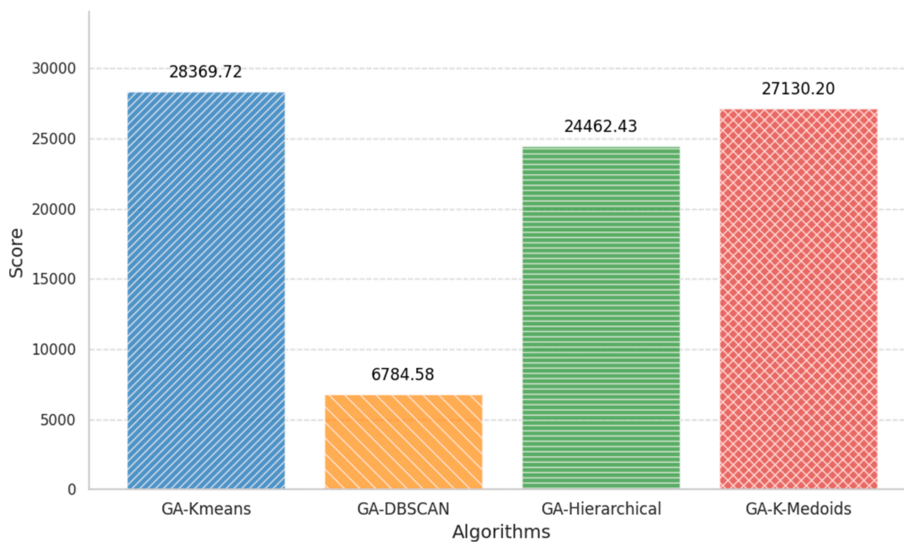

Fig. 14. Comparing Calinski Harabsz Score Gained.

Table 7
Results of Post-Hoc Nemenyi test.

	0	1	2
0	1.000	0.334	0.012
1	0.334	1.000	0.334
2	0.012	0.334	1.000

7.1.2. Energy production comparison

Furthermore, we need to put the centroids in each algorithm into the HOMERPRO and see the results. These are the settings for our simulated HOMERPRO sustainable energy system:

- Solar Panel: Generic Flat Plate PV
- Wind Turbine: Generic 3 kW
- Battery: Generic Generic 1 kWh Li-Ion Battery

The energy production results from both the solar panel and wind turbine are depicted in Figs. 15-19. Evidently, as illustrated in these figures, the solar panel consistently outperforms the wind turbine in generating energy across various clusters and algorithms.

The quality of the clustering findings in terms of compactness and separation was evaluated using the Davies-Bouldin and CH evaluation metrics. In particular, the CH score assisted in assessing cluster dispersion, with higher scores signifying more distinct grouping, while the Davies-Bouldin index sought to reduce intra-cluster similarity while optimizing inter-cluster separation. By using these criteria, it was made sure that the clusters that were created were distinct and appropriate for additional examination. Also, based on the resource information at the centers of each algorithm, we compared the average energy produced by solar panels and wind turbines in 4 algorithms. The evaluation metrics provided a geometric validation of the clustering results, while the average energy production determined which clustering method would perform best for practical implementation in renewable energy optimization. the GA−K-Medoids have the best average energy produced by Wind Turbines; also, the average energy produced by solar panels is relatively high. Therefore, the GA-K-Medoids solution was ultimately chosen based on its superior average energy production.

7.1.3. Cost comparison

The HOMER PRO integration with our optimal locations based on

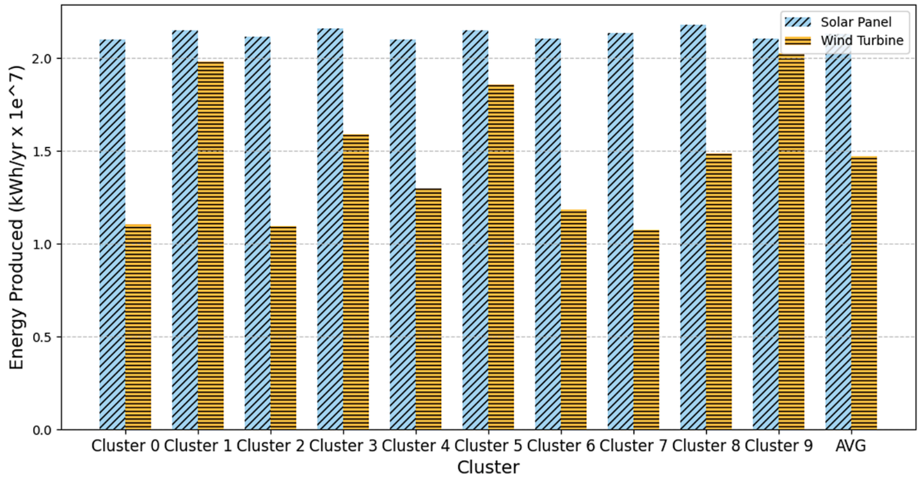

Fig. 15. GA – Kmeans: Energy Produced by Solar Panel and Wind Turbine.

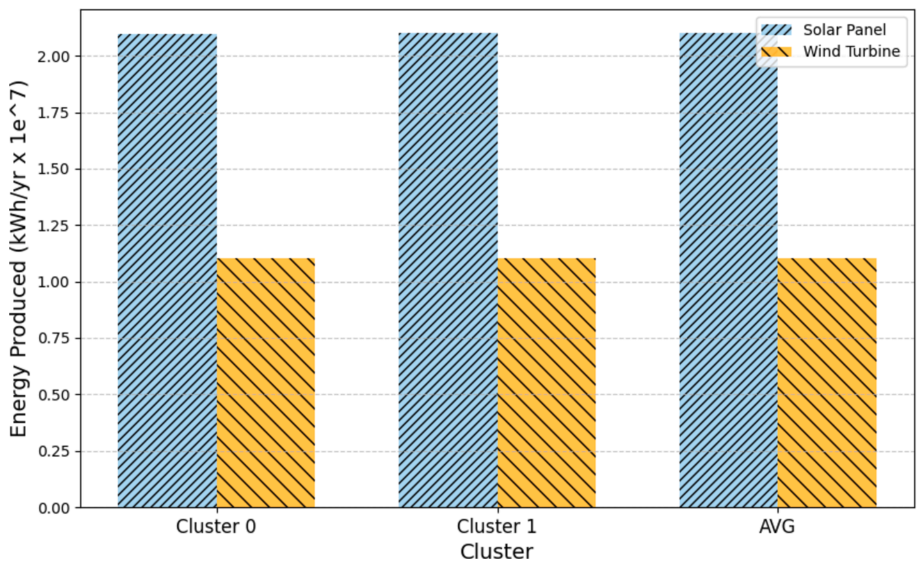

Fig. 16. GA – DBSCAN: Energy Produced by Solar Panel and Wind Turbine.

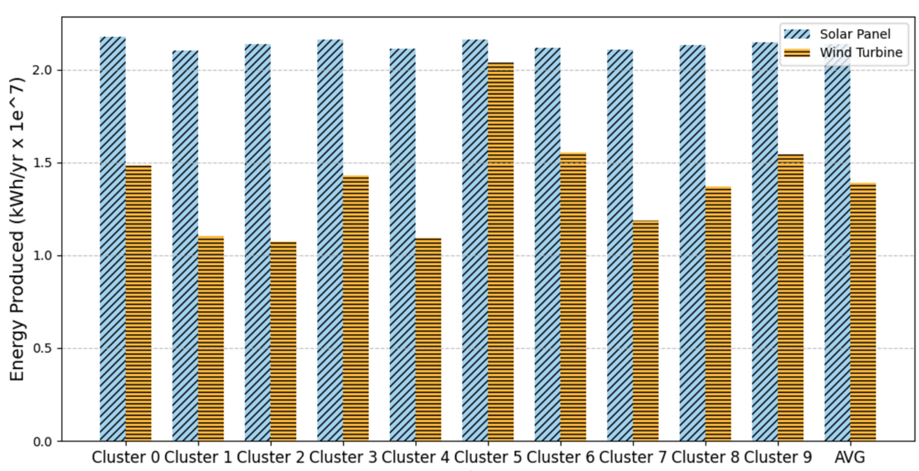

Fig. 17. GA – K – Hierarchical: Energy Produced by Solar Panel and Wind Turbine.

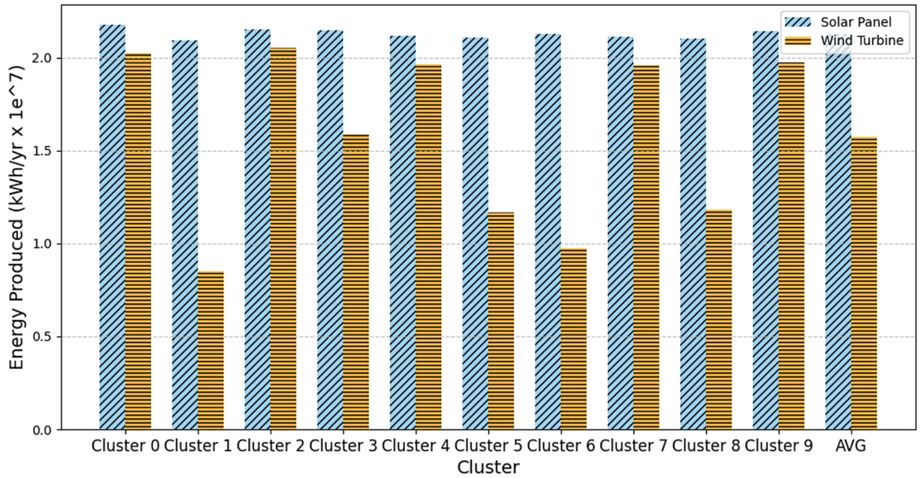

Fig. 18. GA – K – Medoids: Energy Produced by Solar Panel and Wind Turbine.

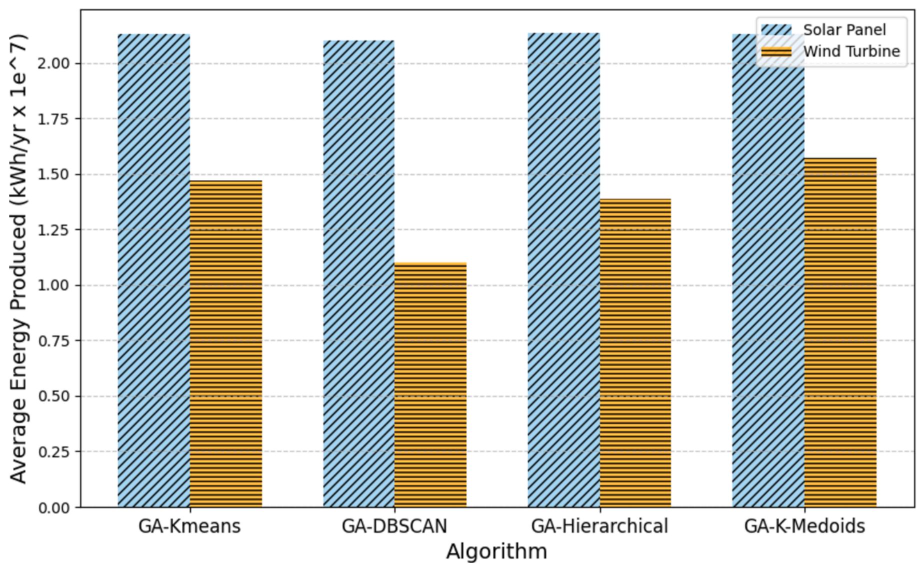

Fig. 19. Average Energy Produced Graph.

collected centroids gave us numerical costs generated by our simulated sustainable energy system (Figs. 20-22). We assessed the costs based on net present cost, levelized cost of energy and operation costs. These metrics are used to illustrate the most optimal costs for implementing a sustainable energy system. These fields are defined further below:

7.1.3.1. Net present cost (NPC). Net Present Costs (NPC) refer to the current value of all expenses incurred by the system throughout its lifespan subtracted from the current value of all the income generated during that same period. These costs involve:

- Capital costs
- Replacement costs
- Operation and Maintenance costs
- Fuel Costs
- Emission Penalties
- The cost associated with purchasing electricity from the utility grid.

Revenues involve:

- Salvage value
- Grid sales

7.1.3.2. Levelized cost of energy (LCOE). The concept of LCOE, or Levelized Cost of Energy, can be described as a metric obtained by dividing the total lifetime costs of a project by the amount of energy it generates. Comparing LCOE enables the assessment of value across the entire project lifespan, making it a valuable tool for making informed decisions about pursuing projects based on economic considerations rather than just utility rates.

7.1.3.3. Operation costs. The operation cost is the value of all costs and revenues, excluding initial capital costs generated annually.

7.1.3.4. Comparison results. A comprehensive analysis of the overall cost metrics shows that while GA K-medoids exhibit the highest energy production (as depicted in Figs. 20-22), they also incur the most substantial financial expenses. It is important to mention that GA-DBSCAN presents challenges in cost comparison due to its fewer centroids. Nevertheless, upon close examination of the figures, it appears that GA-Hierarchical clustering offers the most optimal and cost-effective solution when contrasted with GA-K-means and GA-K-medoids.

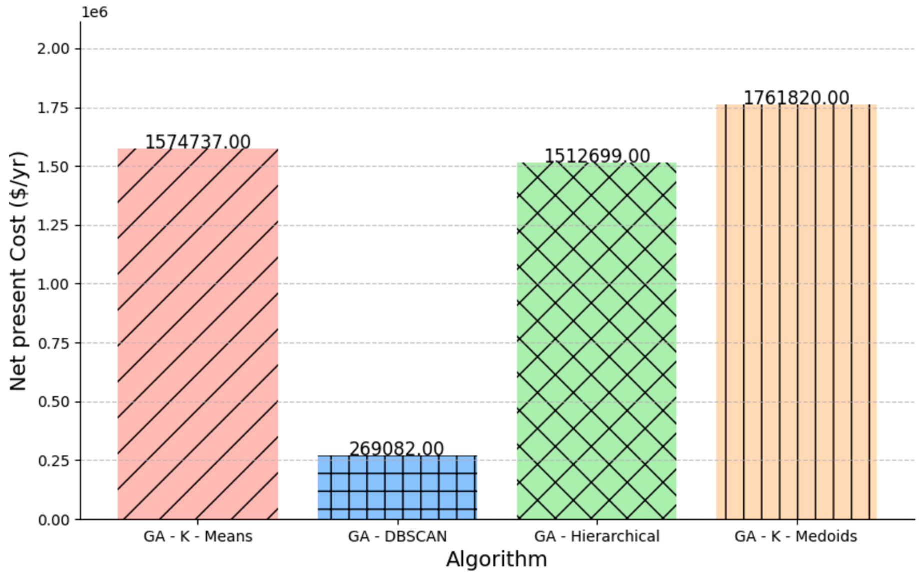

Fig. 20. Net Present Cost Comparison for these four algorithms.

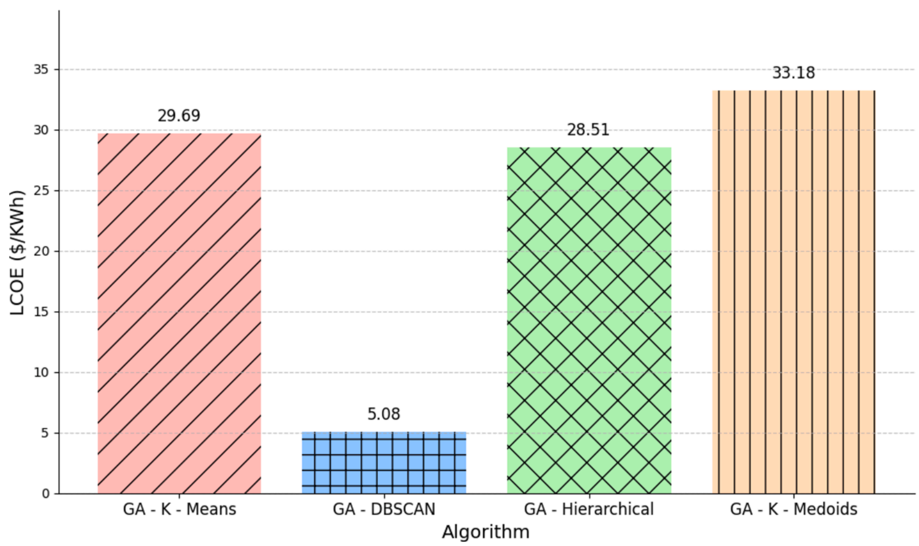

Fig. 21. Levelized Cost of Energy Comparison for these four algorithms.

7.2. Comparison with existing research studies

In comparison to the studies by [2] and [74], this research offers several unique advantages and contributions in the field of renewable energy site selection. While Holloway and his team focused on selecting optimal locations for renewable energy systems in rural regions of Western Australia using K-Means and K-Medoids clustering algorithms, and Khamis et al. [74] targeted remote electrification in Sarawak, Malaysia, through image segmentation and regional techniques, this study provides distinct features and advancements.

Unlike the study by Holloway et al. [2], which primarily utilized K-Means and K-Medoids clustering algorithms, and Khamis et al. [74] who employed image segmentation techniques, our research employs a comprehensive approach integrating various clustering algorithms, including K-Means, DBSCAN, Hierarchical clustering, and K-Medoids. This multi-algorithmic approach allows for a more robust and comprehensive analysis, ensuring the selection of optimal locations for renewable energy stations in rural areas across Australia. Moreover, our study incorporates a genetic algorithm into the clustering process to identify the most appropriate clustering method dynamically. This iterative process enhances the efficiency and effectiveness of our site selection methodology by adapting to the characteristics of the dataset and optimizing the clustering results.

While the previous two studies focused on specific regions, this study provides insights and techniques applicable to rural areas across Australia as a whole. This is because a more extensive and reliable dataset covering all Australian towns has been found. This broader scope enhances the universality and applicability of our research findings, making them relevant to a wider range of stakeholders involved in promoting renewable energy development in Australia.

This research integrates HOMER Pro software to estimate the solar and wind energy potential for each identified location. This integration of energy output data enhances the accuracy and reliability of our site selection process, providing valuable insights for evaluating the efficacy of different clustering algorithms.

Overall, this study contributes to the field of renewable energy site selection by presenting a comprehensive methodology, leveraging multiple clustering algorithms, integrating a genetic algorithm for

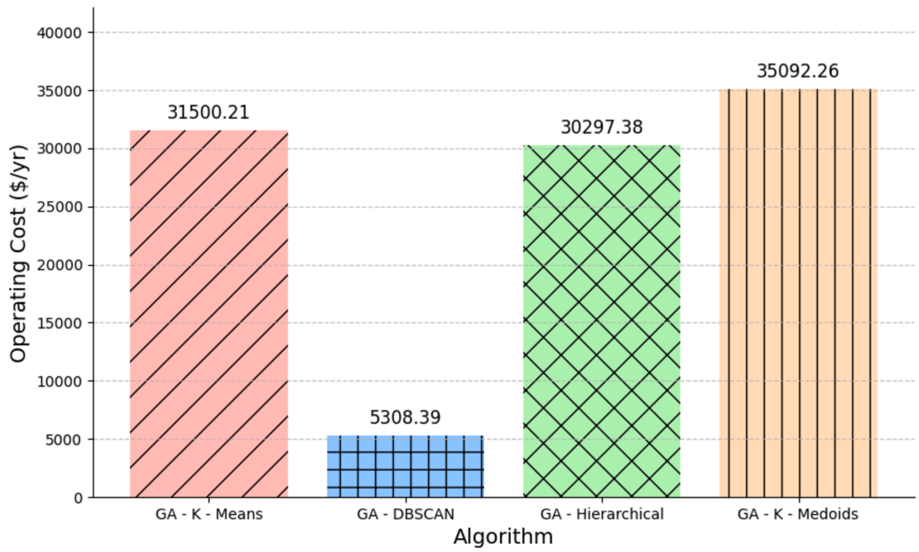

Fig. 22. Operating Cost for these four algorithms.

dynamic optimization, and providing insights applicable to rural areas across Australia. These advancements offer significant value for informing future research and policy decisions aimed at promoting renewable energy development in Australia.

Based on the gained results from this study, there exist some directions for improvement of the current study. For example, in future research, it would be valuable to explore the integration of economic feasibility studies with clustering methodologies to assess the cost-effectiveness of renewable energy projects comprehensively. The decision-making process may also be improved by utilizing cutting-edge technology like machine learning for real-time data analysis and predictive analytics. Moreover, a comprehensive viewpoint for sustainable implementation would be suggested by assessing the socioeconomic effects, taking into account the advantages and involvement of the local population. Furthermore, planning for resilience could be strengthened by examining how hybrid renewable energy systems operate in a variety of climatic circumstances, including extreme weather events.

7.3. Limitations

Homer Pro is a recognized tool in the field of renewable energy analysis, widely used for its robust simulation capabilities. It enables detailed modeling of different energy sources by considering varying inputs like weather data, system configurations, and operational strategies. The choice of Homer Pro was guided by its extensive use in preliminary assessments of renewable projects, where direct measurement data may not be readily available. However, it is recognized that solely relying on simulation data can introduce uncertainties. Although Homer Pro has a detailed economic calculation, the detailed calculation (mathematical modeling) is not revealed, and it acts as a black box with limited flexibility in changing the input data. This limitation can be considered as future work.

Furthermore, this research aims to address the challenges of site selection for renewable energy sites through a technical and data-driven approach to minimize adverse environmental impacts. However, it is recognized that establishing these sites in rural areas may indeed impact local ecosystems and biodiversity. Therefore, it is encouraged as future research and practice to conduct comprehensive environmental assessments and implement protective measures when siting renewable energy projects to minimize adverse effects on ecosystems and actively promote sustainable development.

8. Conclusion

This research highlights the efficacy of clustering algorithms combined with genetic optimization in identifying optimal locations for HRESs in rural Australia. The study effectively identified appropriate regions using the K-Means, DBSCAN, Hierarchical, and K-Medoids algorithms; clustering efficiency was further improved by genetic algorithms. A thorough assessment of the solar and wind energy potential at cluster centers was made possible by integration with HOMER Pro software. Based on metrics such as the Silhouette score, the genetic K-Means method was found to be the most computationally cost-effective approach, while the GA K-Medoids algorithm produced the highest average energy output of 33.18 kWh/year, outperforming other methods.This information is useful for the development of renewable energy.

Also, it is worthy to mention that the study includes limitations that offer potential for further investigation despite its contributions. Interestingly, the analysis concentrated on finding resource-rich areas rather than taking into consideration the whole cost of building renewable energy installations. Although HOMER software is useful for energy simulation, however its closed-system design and restricted input flexibility leads to some difficulties. In order to take into account wider economic and environmental factors, future study might minimize installation and operating costs, include thorough cost evaluations, and improve techniques.

Moreover, for the aim of increasing, future developments could investigate different distance metrics, including the Haversine formula. Adding more characteristics, such area and elevation, could provide more detailed information about how geographic characteristics affect the design of energy systems. By improving hybrid renewable systems' scalability and application, these initiatives hope to result in sustainable energy development.

CRediT authorship contribution statement

Iman Rahimi: Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Conceptualization. **Mufei Li:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation. **James Choon:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation. **Dane Pamuspusan:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Yinpeng Huang:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation. **Binzhen He:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation. **Alan Cai:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation. **Mohammad Reza Nikoo:** Writing – review & editing, Visualization. **Amir H. Gandomi:** Writing – review & editing, Supervision, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecmx.2024.100855.

Data availability

Data will be made available on request.

References

[1] Vine E. The integration of energy efficiency. Renewable Energy, DemandResponse and Climate Change: Challenges and Opportunities for Evaluatorsand Planners 2007.
[2] Holloway R, Ho D, Delotavo C, Xie WY, Rahimi I, Nikoo MR, et al. Optimal location selection for a distributed hybrid renewable energy system in rural Western Australia: a data mining approach. Energ Strat Rev 2023;50:101205.
[3] Nematian J, Rahimi I. Feasibility study of using renewable energies in Iranian Seas: a comparative study. Renew Energy 2022;189:383–91.
[4] Teske S, Fattal A, Lins C, Hullin M, Williamson LE. Renewables global futures report: Great debates towards 100% renewable energy, 2017.
[5] Alamdari P, Nematollahi O, Alemrajabi AA. Solar energy potentials in Iran: a review. Renew Sustain Energy Rev 2013;21:778–88.
[6] Fazelpour F, Soltani N, Rosen MA. Feasibility of satisfying electrical energy needs with hybrid systems for a medium-size hotel on Kish Island Iran. Energy 2014;73: 856–65. https://doi.org/10.1016/j.energy.2014.06.097.
[7] Pérez-López P, Gschwind B, Blanc P, Frischknecht R, Stolz P, Durand Y, et al. ENVI-PV: an interactive Web Client for multi-criteria life cycle assessment of photovoltaic systems worldwide. Prog Photovolt Res Appl 2017;25(7):484–98. https://doi.org/10.1002/pip.2841.
[8] Zahedi A. Australian renewable energy progress. Renew Sustain Energy Rev 2010; 14(8):2208–13.
[9] Byrne J, Kurdgelashvili L, Mathai M, Kumar A, Yu J, Zhang X, et al. World solar energy review: technology, markets and policies. Center for Energy and Environmental Policies Report; 2010.
[10] Ellabban O, Abu-Rub H, Blaabjerg F. Renewable energy resources: current status, future prospects and their enabling technology. Renew Sustain Energy Rev 2014; 39:748–64.
[11] Sawin JL, Martinot E. Renewables 2013: Global status report. REN21, 2013.
[12] Bahadori A, Nwaoha C. A review on solar energy utilisation in Australia. Renew Sustain Energy Rev 2013;18:1–5.
[13] Shafiullah GM, Amanullah MTO, Ali ABMS, Jarvis D, Wolfs P. Prospects of renewable energy–a feasibility study in the Australian context. Renew Energy 2012;39(1):183–97.
[14] Hemer MA, Manasseh R, McInnes KL, Penesis I, Pitman T. Perspectives on a way forward for ocean renewable energy in Australia. Renew Energy 2018;127:733–45.
[15] Li HX, Edwards DJ, Hosseini MR, Costin GP. A review on renewable energy transition in Australia: an updated depiction. J Clean Prod 2020;242:118475.
[16] Li J, Liu P, Li Z. Optimal design and techno-economic analysis of a solar-wind-biomass off-grid hybrid power system for remote rural electrification: a case study of west China. Energy 2020;208:118387.
[17] Martin NJ, Rice JL. Developing renewable energy supply in Queensland, Australia: a study of the barriers, targets, policies and actions. Renew Energy 2012;44: 119–27.
[18] Yusaf T, Goh S, Borserio JA. Potential of renewable energy alternatives in Australia. Renew Sustain Energy Rev 2011;15(5):2214–21.
[19] Elkadeem MR, Younes A, Sharshir SW, Campana PE, Wang S. Sustainable siting and design optimization of hybrid renewable energy system: a geospatial multi-criteria analysis. Appl Energy 2021;295:117071.
[20] Rahimi I, Nematian J. A bibliometric analysis on optimization solution methods applied to supply chain of solar energy. Arch Comput Meth Eng 2022:1–19.
[21] Rahimi I, Nikoo MR, Gandomi AH. Techno-economic analysis for using hybrid wind and solar energies in Australia. Energ Strat Rev 2023;47:101092.
[22] Elkadeem MR, Younes A, Mazzeo D, Jurasz J, Campana PE, Sharshir SW, et al. Geospatial-assisted multi-criterion analysis of solar and wind power geographical-technical-economic potential assessment. Appl Energy 2022;322:119532.
[23] Hosseinifard F, Hosseinpour M, Salimi M, Amidpour M. Achieving net zero energy penalty in post-combustion carbon capture through solar Energy: parabolic trough and photovoltaic technologies. Energy Convers Manage: X 2024;24:100757.
[24] Rekik S, El Alimi S. Optimal wind-solar site selection using a GIS-AHP based approach: a case of Tunisia. Energy Convers Manage: X 2023;18:100355.
[25] Ajanovic A, Haas R, Nakicenovic N. Economic analysis of production and use of hydrogen from solar energy, wind, hydropower and biomass. ISES Solar World Congress 2007, ISES 2007, 4, 2496–2500.
[26] Herbert GMJ, Iniyan S, Sreevalsan E, Rajapandian S. A review of wind energy technologies. Renew Sustain Energy Rev 2007;11(6):1117–45.
[27] Wagner H-J. Introduction to wind energy systems. EPJ Web of Conferences 2018; 189:5.
[28] Chambon CL, Karia T, Sandwell P, Hallett JP. Techno-economic assessment of biomass gasification-based mini-grids for productive energy applications: the case of rural India. Renew Energy 2020;154:432–44.
[29] Jahangir MH, Cheraghi R. Economic and environmental assessment of solar-wind-biomass hybrid renewable energy system supplying rural settlement load. Sustain Energy Technol Assess 2020;42:100895.
[30] Kaundinya DP, Balachandra P, Ravindranath NH, Ashok V. A GIS (geographical information system)-based spatial data mining approach for optimal location and capacity planning of distributed biomass power generation facilities: a case study of Tumkur district, India. Energy 2013;52:77–88.
[31] Babaei M, Abazari A, Soleymani MM, Ghafouri M, Muyeen SM, Beheshti MTH. A data-mining based optimal demand response program for smart home with energy storages and electric vehicles. J Storage Mater 2021;36:102407.
[32] Gawusu S, Mensah RA, Das O. Exploring distributed energy generation for sustainable development: a data mining approach. J Storage Mater 2022;48: 104018.
[33] LLC HE. HOMER (Hybrid Optimization of Multiple Energy Resources) Software. Version.
[34] Lydia EL, Kumar BP, Ramya D. Generation of dynamic energy management using data mining techniques basing on big data analytics isssues in smart grids. Int J Eng Technol 2018;7(2.26):85–9.
[35] Bahramara S, Moghaddam MP, Haghifam MR. Optimal planning of hybrid renewable energy systems using HOMER: a review. Renew Sustain Energy Rev 2016;62:609–20.
[36] Kazak JK, Mrówczyńska M, Skiba M, Bazan-Krzywoszańska A, Świader M, Tokarczyk-Dorociak K, et al. A decision support system for the planning of hybrid renewable energy technologies. IOP Conf Ser: Earth Environ Sci 2021;701(1): 12012.
[37] Kanata S, Baqaruzi S, Muhtar A, Prasetyawan P, Winata T. Optimal planning of hybrid renewable energy system using Homer in Sebesi Island, Indonesia. Int J Renew Energ Res (IJRER) 2021;11(4):1507–16.
[38] Jbaihi O, Ouchani F, Merrouni AA, Cherkaoui M, Ghennioui A, Maaroufi M. An AHP-GIS based site suitability analysis for integrating large-scale hybrid CSP+ PV plants in Morocco: an approach to address the intermittency of solar energy. J Clean Prod 2022;369:133250.
[39] Masoomi B, Sahebi IG, Fathi M, Yldrm F, Ghorbani S. Strategic supplier selection for renewable energy supply chain under green capabilities (fuzzy BWM-WASPAS-COPRAS approach). Energ Strat Rev 2022;40:100815.
[40] Narayanamoorthy S, Parthasarathy TN, Pragathi S, Shanmugam P, Baleanu D, Ahmadian A, et al. The novel augmented Fermatean MCDM perspectives for identifying the optimal renewable energy power plant location. Sustain Energy Technol Assess 2022;53:102488.
[41] Demolli H, Dokuz AS, Ecemis A, Gokcek M. Location-based optimal sizing of hybrid renewable energy systems using deterministic and heuristic algorithms. Int J Energy Res 2021;45(11):16155–75.
[42] de Barros Franco DG, Steiner MTA. Clustering of solar energy facilities using a hybrid fuzzy c-means algorithm initialized by metaheuristics. J Clean Prod 2018; 191:445–57.
[43] Wang Q, Yang X. Investigating the sustainability of renewable energy–an empirical analysis of European Union countries using a hybrid of projection pursuit fuzzy clustering model and accelerated genetic algorithm based on real coding. J Clean Prod 2020;268:121940.
[44] Biboum AC, Yilanci A, Mouangue R. Comparative analysis of hybrid renewable energy systems based on concentrating solar and biomass technologies for Faro-Poli. Cameroon Environmental Progress & Sustainable Energy 2023;42(1):e13933.
[45] Menesy AS, Sultan HM, Habiballah IO, Masrur H, Khan KR, Khalid M. Optimal configuration of a hybrid photovoltaic/wind turbine/biomass/hydro-pumped storage-based energy system using a heap-based optimization algorithm. Energies 2023;16(9):3648.
[46] Marocco P, Ferrero D, Lanzini A, Santarelli M. The role of hydrogen in the optimal design of off-grid hybrid renewable energy systems. J Storage Mater 2022;46: 103893.
[47] Di Grazia S, Tina GM. Optimal site selection for floating photovoltaic systems based on Geographic Information Systems (GIS) and Multi-Criteria Decision Analysis (MCDA): a case study. Int J Sustain Energ 2023:1–23.
[48] Wei P, Bamisile O, Adun H, Cai D, Obiora S, Li J, et al. Bibliographical progress in hybrid renewable energy systems' integration, modelling, optimization, and artificial intelligence applications: a critical review and future research perspective. Energy Sources Part A 2023;45(1):2058–88.
[49] Kumar R, Channi HK. A PV-Biomass off-grid hybrid renewable energy system (HRES) for rural electrification: Design, optimization and techno-economic-environmental analysis. J Cleaner Prod 2022;349:131347.
[50] Alavi SMS, Maleki A, Noroozian A, Khaleghi A. Simultaneous optimal site selection and sizing of a grid-independent hybrid wind/hydrogen system using a hybrid

optimization method based on ELECTRE: a case study in Iran. Int J Hydrogen Energy 2024;55:970–83.
[51] Heydari F, Maleki A, Jabari Moghadam A, Haghighat S. Emplacement of the photovoltaic water pumping system in remote areas by a multi-criteria decision-making method: a case study. Front Energy Res 2021;9:770981.
[52] Shahraki Shahdabadi R, Maleki A, Haghighat S, Ghalandari M. Using multi-criteria decision-making methods to select the best location for the construction of a biomass power plant in Iran. J Therm Anal Calorim 2021;145:2105–22.
[53] Sadeghi A, Maleki A, Haghighat S. Techno-economic analysis and optimization of a hybrid solar-wind-biomass-battery framework for the electrification of a remote area: a case study. Energy Convers Manage: X 2024;24:100732.
[54] Batista NE, Carvalho PCM, Fernández-Ramirez LM, Braga APS. Optimizing methodologies of hybrid renewable energy systems powered reverse osmosis plants. Renew Sustain Energy Rev 2023;182:113377.
[55] Seyed Alavi SM, Maleki A, Khaleghi A. Optimal site selection for wind power plant using multi-criteria decision-making methods: a case study in eastern Iran. International Journal of Low-Carbon Technologies 2022;17:1319–37.
[56] Maleki A. Optimization based on modified swarm intelligence techniques for a stand-alone hybrid photovoltaic/diesel/battery system. Sustainable Energy Technol Assess 2022;51:101856.
[57] Maleki A. Optimal operation of a grid-connected fuel cell based combined heat and power systems using particle swarm optimisation for residential sector. Int J Ambient Energy 2021;42(5):550–7.
[58] Lazzari F, Mor G, Cipriano J, Solsona F, Chemisana D, Guericke D. Optimizing planning and operation of renewable energy communities with genetic algorithms. Appl Energy 2023;338:120906.
[59] Sousa J, Lagarto J, Camus C, Viveiros C, Barata F, Silva P, et al. Renewable energy communities optimal design supported by an optimization model for investment in PV/wind capacity and renewable electricity sharing. Energy 2023;283:128464.
[60] Javed MS, Ma T, Jurasz J, Canales FA, Lin S, Ahmed S, et al. Economic analysis and optimization of a renewable energy based power supply system with different energy storages for a remote island. Renew Energy 2021;164:1376–94.
[61] Byrnes L, Brown C, Wagner L, Foster J. Reviewing the viability of renewable energy in community electrification: the case of remote Western Australian communities. Renew Sustain Energy Rev 2016;59:470–81.
[62] Clifton J, Boruff BJ. Assessing the potential for concentrated solar power development in rural Australia. Energy Policy 2010;38(9):5272–80.
[63] Fornarelli R, Shahnia F, Anda M, Bahri PA, Ho G. Selecting an economically suitable and sustainable solution for a renewable energy-powered water desalination system: a rural Australian case study. Desalination 2018;435:128–39.
[64] Hicks J, Ison N. Community-owned renewable energy (CRE): opportunities for rural Australia. Rural Soc 2011;20(3):244–55.
[65] McHenry MP. Remote area power supply system technologies in Western Australia: New developments in 30 years of slow progress. Renew Energy 2009;34(5): 1348–53.
[66] Yiridoe EK, et al. Social acceptance of wind energy development and planning in rural communities of Australia: a consumer analysis. Energy Policy 2014;74: 262–70.
[67] Maulik U, Bandyopadhyay S. Performance evaluation of some clustering algorithms and validity indices. IEEE Trans Pattern Anal Mach Intell 2002;24(12): 1650–4.
[68] Caliński T, Harabasz J. A dendrite method for cluster analysis. Communications in Statistics-Theory and Methods 1974;3(1):1–27.
[69] Aranganayagi S, Thangavel K. Clustering categorical data using silhouette coefficient as a relocating measure. International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007), 2, 13–17.
[70] Kumar P, Pukale R, Kumabhar N, Patil U. Optimal design configuration using HOMER. Procedia Technol 2016;24:499–504.
[71] Al-Karaghouli A, Kazmerski LL. Optimization and life-cycle cost of health clinic PV system for a rural area in southern Iraq using HOMER software. Sol Energy 2010; 84(4):710–4.
[72] Shahzad MK, Zahid A, ur Rashid T, Rehan MA, Ali M, Ahmad M. Techno-economic feasibility analysis of a solar-biomass off grid system for the electrification of remote rural areas in Pakistan using HOMER software. Renew Energy 2017;106: 264–73.
[73] Jahangir MH, Mousavi SA, Rad MAV. A techno-economic comparison of a photovoltaic/thermal organic Rankine cycle with several renewable hybrid systems for a residential area in Rayen Iran. Energy Convers Manag 2019;195: 244–61.
[74] Khamis A, Khatib T, Yosliza NAHM, Azmi AN. Optimal selection of renewable energy installation site in remote areas using segmentation and regional technique: a case study of Sarawak Malaysia. Sustainable Energy Technologies and Assessments 2020;42:100858.