Engineering Applications of Artificial Intelligence 126 (2023) 106848

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Smart investigation of artificial intelligence in renewable energy system technologies by natural language processing: Insightful pattern for decision-makers

Kamran Niroomand [a], Noori M. Cata Saady [b,*], Carlos Bazan [c], Sohrab Zendehboudi [d], Amilcar Soares [e], Talib M. Albayati [f]

[a] Department of Computer Engineering, Memorial University, St. John's, NL A1B 3X5, Canada
[b] Department of Civil Engineering, Memorial University, St. John's, NL A1B 3X5, Canada
[c] Faculty of Business, Memorial University, St. John's, NL A1B 3X5, Canada
[d] Department of Process Engineering, Memorial University, St. John's, NL A1B 3X5, Canada
[e] Department of Computer Science, Memorial University, St. John's, NL A1B 3X5, Canada
[f] Chemical Engineering Department, University of Technology, Baghdad 10071, Iraq

ARTICLE INFO

Keywords:
Natural language processing
Artificial intelligence
Text mining
Topic modeling
Pattern identification
Renewable energy

ABSTRACT

This study aims to provide a framework which enables decision-makers and researchers to identify AI technology patterns in renewable energy systems from a massive data set of textual data. However, the study was challenged by the Scopus database limitation that allows users to retrieve only 2000 documents per query. Therefore, we developed a search engine based on the Scopus Application Programming Interface (API) that enables us to download an unlimited number of documents per query based on our desirable settings. We extracted 5661 renewable energy systems-related publications from Scopus database and leveraged Natural Language Processing (NLP) and unsupervised algorithms to identify the most frequent computational science models and dense meta-topics and investigate their evolution throughout the period 2000-2021. Our findings showed 7 meta-topics based on the class-based Term Frequency-Inverse Document Frequency (c-TD-IDF) score and term score decline graph. Emerging advanced algorithms, such as different deep learning architectures, directly impacted growing meta-topics involving problems with uncertainty and dynamic conditions.

1. Introduction

With the ever-increasing demand for energy, limitations of fossil fuel resources, and concerns about sustainability, renewable energy systems are increasingly gaining attention from governments, businesses, and research institutes worldwide (Azarpour et al., 2022; Tazikeh et al., 2022). Hence, developing a clear technology roadmap is important to integrate science and technology with business planning meaningfully based on medium to long-term market direction and goals (Amer and Daim, 2010). Performing intelligence investigation on upcoming technologies and clear technology roadmaps will assist governments and industries in making smart investing decisions and maintaining their competitive edge (Angelo et al., 2017). This study focuses on the research direction of applying AI text modeling techniques for identifying the most common strategies employed by renewable energy systems literature, aiming at establishing a technology selection and research perspective in this domain. The significant growth of the application of AI modeling in the renewable energy domain creates a massive amount of data and information within research papers, registered patents, and reports.

Based on the literature, several research studies have been conducted in different engineering and science fields with focus on AI applications for prediction and classification purposes (Zendehboudi et al., 2018; Seyyedattar et al., 2020). In this regard, some researchers have focused on leveraging AI in different renewable energy systems, from electrical and power systems (Zhou et al., 2022; Zhang et al., 2023) to biogas systems. Tufaner and Demirci (2020) employed a three-layer artificial neural network on lab-scale data to forecast biogas production rate by considering different features such as effluent alkalinity, organic loading rate, and effluent chemical oxygen demand. Chiu et al. (2022) applied a hybrid machine learning model, random forest, and long short-term memory by analyzing important feeds for biogas production for optimization based on a biogas plant dataset in China. They gained significant performance without conducting intensive feature engineering. Table 1 presents recently published similar

* Corresponding author.
E-mail addresses: skniroomand@mun.ca (K. Niroomand), nsaady@mun.ca (N.M.C. Saady), cbazan@mun.ca (C. Bazan), szendehboudi@mun.ca (S. Zendehboudi), amilcarsj@mun.ca (A. Soares), Talib.M.Naieff@uotechnology.edu.iq (T.M. Albayati).

https://doi.org/10.1016/j.engappai.2023.106848
Received 27 August 2022; Received in revised form 25 June 2023; Accepted 19 July 2023
Available online 2 August 2023

Table 1
Previous scientific works at the intersection of AI and biogas.

Objective	Model(s)	Result	Reference
Optimizing biogas purification	DNN	The optimum ranges of: C/N (15.04–18.95), BOD/COD (0.763–0.818), TS (8.1–10.6%) and T.VS (38.19–49.46%). Large BOD/COD impacts biogas purification. pH > 7 can improve biogas purification.	Mahmoodi-Eshkaftaki and Ebrahimi (2021)
Predicting biogas production rate of food waste dry anaerobic digestion considering HRT, SRT, soluble chemical oxygen demand, total VFA, total and ammonia features.	RNN	Solid retention time and water content are important features in biogas reaction. Increasing intermediate materials, like VFAs, were easily converted into methane at higher water contents.	Seo et al. (2021)
Modeling chemical processes within biogas production system	SNN	Considering ten days data points, model can predict chemical process up to the 100th day with a significant accuracy based on lab-scale data.	Capizzi et al. (2020)
Predicting biogas production of fruits and vegetables waste considering different operational parameters	ANN	Predicted the performance with 85% accuracy.	Gonçalves Neto et al. (2021)
Identifying important operational parameters and predicting the biogas system production rate	RF KNN SVM GLMNET	Total carbon was identified as a most important feature. KNN performed well in the regression task with RMSE of 26.6 and logistic regression multiclass model gained accuracy of 73%.	Wang et al. (2020)
Predicting biogas systems performance of vegetables, fruits waste	ANFIS LSSVM	LSSVM performed better than ANFIS. LSSVM had MRE % and MSE of 2.951 and 0.0001 respectively, compared to 29.318 and 0.0039 for ANFIS.	Yang et al. (2021)
Predicting methane production in a biogas plant	Gompertz ML HML and Gompertz	HML was the best in predicting next-day biogas production and reduced the error by 53%. MAPE of HM (4.52%) < ML (4.84%) < the Gompertz model (9.61%).	Hansen et al. (2020)
Building a predictive model of biogas yield and establishing optimal conditions for cow manure and maize straw biogas process	ANFIS	$R^2 = 0.99$ and the model suggested conditions increased the production by 8%.	Zareei and Khodaei (2017)
Predicting biogas production of spent mushroom compost in thermophilic and mesophilic laboratory conditions	ANN ANFIS	RMSE and R^2 in mesophilic condition: ANFIS are 0.1940 and 0.9998, ANN are 0.780 and 0.9981, and logistic model are 0.5111 and 0.9992, respectively. In thermophilic condition, the values of RMSE and R^2 were indicated as 0.3033 and 0.9997 for ANFIS, 0.3430 and 0.9992 for ANN, and 0.5506 and 0.9991 for the logistic model, respectively.	Najafi and Faizollahzadeh Ardabili, 2018
Predicting biogas production rate based on industrial data and finding important operation parameters	ACO GA ANN	$R^2 = 0.9$ and prediction error = 6.24%.	Beltramo et al. (2019)

Note: DNN = Deep neural networks; RNN = Recurrent neural network; SNN = Spiking neural network; ANN = Artificial neural networks; RF = Random forest; KNN = K- nearest neighbor; RMSE = Root mean square error; mean relative error = MRE; Mean squared error = MSE; Mean absolute percentage error = MAPE; SVM = Support vector machine; GLMNET = Generalized linear models fitting package via penalized maximum likelihood; LSSVM = Least square support vector machine; ML = Machine learning; HML = Hybrid of machine learning and Gompertz; ANFIS = Adaptive neuro-fuzzy inference system (ANFIS); ACO = Ant colony optimization (ACO); GA = Genetic algorithms (GA).

research leveraging various AI models for optimizing and forecasting biogas systems' performance.

Like other domains, a problem that investors and decision-makers in governmental and industrial sectors face is that they can get confused among various generated information, specifically in the field of AI with a wide variety of techniques and approaches. Developing an almost automatic framework that can be employed to address the mentioned problem is crucial. More specifically, such a method can provide accurate insights quickly for various research domains, even complex areas, to empower decision-makers to have a better perception of the research dynamics and assist them in determining research and development (R&D) strategies. This study aimed to build a research landscape and give decision-makers and researchers in related domains an accurate insight into implementing AI algorithms in renewable energy systems from a scientific standpoint. There is a high level of correspondence between scientific papers and patents; in other words, patent quality can be measured by academic papers referenced (Poege et al., 2019; Coupé, 2003). Besides, the World Intellectual Property Organization specified that more than 90% of inventions are observed in patent papers (Souili et al., 2015). Hence, texts of scientific papers are beneficial resources for investigating technology development throughout a timeline, and their understanding will benefit research and development. However, analyzing the vast textual literature and data is time-consuming and prone to mistakes. A powerful solution for this problem is using NLP techniques (Chowdhary, 2020). The NLP techniques have various applications such as information extraction, automated text summarization, question answering system, and speech recognition (Jusoh, 2018). Topic modeling is an unsupervised strategy that is a subfield of NLP, which is a valuable technique for achieving a high level of understanding of a large amount of unstructured text data (Hannigan et al., 2019). The basis of NLP is considering co-occurrences of a word in similar corpora (Daenekindt and Huisman, 2020). To clarify, co-occurrence rules empower machine to discover and group related concepts within the set of documents or records.

This idea implies that when concepts are often found together in documents and records, that co-occurrence illustrates a hidden relationship that is probably of value in categorizing definitions. Various topic modeling methods have emerged in previous years, such as Latent Dirichlet Allocation (LDA), the most frequently leveraged algorithm in topic modeling (Jockers and Thalken, 2020). Also, there are other less frequent methods like Bidirectional Encoder Representations from Transformers Topic (BERTopic) (Grootendorst, 2022), Top2Vec (Angelov, 2020), Structural Topic Modelling (STA) (Lindstedt, 2019), Correlation Explanation (CorEX) (Gallagher et al., 2017), Non-Negative Matrix Factorisation (NMF) (Xu et al., 2003), and Latent Semantic Analysis (LSA) (Landauer et al., 1998).

Previous research applied NLP models to scientific papers for topic modeling and trend identification. Mosallaie et al. (2021) employed NLP techniques to investigate the application of AI in scientific papers on cancer-related domains. Tran et al. (2019) employed LDA in scientific papers between 1991 and 2018 to perform topic modeling and provided static insight into the pattern of AI in the cancer domain. Jallan et al. (2019) applied the LDA model to detect current patterns in "construction-defect litigation cases". Lee et al. (2019) performed an accurate NLP-based model to extract contract risk, systematically detecting "poisonous" terms to help related companies manage their contracts. It performed well, achieving 81.8% area under the precise recall curve. Other studies utilized NLP approaches for technology forecasting. Kyebambe et al. (2017) clustered similar technologies considering patent characteristics and predicted new technologies one year forward. Lee et al. (2019) investigated the value of patents by leveraging feed-forward artificial neural networks and conducted an evaluation analysis system for emerging technologies. Johri and Olds (2011) applied topic modeling techniques to investigate emerging topics in engineering education between 2000 and 2008. More specifically, they leveraged LDA, an unsupervised learning method (Blei et al., 2003), as a topic modeling method to extract topics and their top 20 keywords correspondence in engineering education. They also extracted key phrases alongside their corresponding frequency values to quantitatively analyze their trends over time. Based on their results, some topics, like the "global and interaction aspect of engineering education", observed a considerable increase, while other topics remained almost constant over the period. Moreover, topic modeling has various applications in bioinformatic. For instance, Chen et al. (2011) investigated the abundance data pertaining to microbial-community taxa. In their research, they examined protein-coding sequences and their corresponding taxonomical levels as provided by NCBI. They introduced an enhanced version of LDA model known as LDA-B, which incorporated a background distribution consisting of commonly shared functional elements. This extension aimed to uncover functional groups within the dataset. The LDA-B model was employed to identify these functional groups. The researchers utilized the genome set as the document corpus, which encompassed a combination of various functional groups. Each functional group, or topic, represented a weighted combination of functional elements, analogous to the concept of "words" in traditional language models. Li et al. (2022) presents a novel method for semantic recognition of ship motion patterns during port entry and departure. The proposed approach utilizes LDA model, enabling the identification of these patterns from vast amounts of trajectory data while enhancing the interpretability of the outcomes. The pipeline consists of three main modules: trajectory preprocessing, semantic processing, and knowledge discovery. They considered topic coherence and topic correlation metrics to find out the optimal number of topics. In addition, a visualization platform was designed based on ArcGIS and Electronic Navigational Charts, facilitating the analysis of ship motion pattern knowledge. Finally, the proposed method was validated considering data from the Tianjin port in northern China. The experimental results demonstrate the method's efficacy in identifying 17 representative inbound and outbound motion patterns from automatic identification system data and uncovering detailed ship motion characteristics within each pattern.

There are other researchers that leveraged topic modeling in other domains like transportation (Sun and Yin, 2017), hydropower (Jiang et al., 2016), communication research (Maier et al., 2018), smart factory (Yang et al., 2018), and marketing (Reisenbichler and Reutterer, 2019).

To the best of our knowledge, this study is the first to comprehensively investigate the trend of AI topics in the renewable energy systems domain using NLP techniques. Although some studies appeared on trend analysis with limited scope to only one area, such as solar energy or anaerobic digestion, they did not employ NLP techniques (Dong et al., 2012; Ren et al., 2018). We built an automated pipeline which can retrieve unlimited number of documents and extract their hidden information.

2. Methodology

Fig. 1 presents the methodology employed in this study. First, a dataset was built by collecting raw data from Scopus. This raw dataset contains all renewable energies scientific papers in which AI modeling has been used throughout the period 2000–2021. The collected dataset was preprocessed within three steps specified in the "Preprocessing dataset" section. The next step is conducting different exploratory analyses on the preprocessed dataset. The preprocessed dataset was used as the input for the BERTopic model to generate topics. Finally, the created topics were merged into dense meta-topics by domain experts, and their evolution was investigated over time by the dynamic topic modeling (DTM) method.

2.1. Collecting raw dataset

We chose Scopus since it is one of the most comprehensive database for published papers. The Scopus database contains more than 22000 journals and books in renewable energy and computer science sectors, more specifically, 335 journals and 6699 books for renewable energies, and 1337 journals and 14 111 books for computer science. The dataset was extracted from the Scopus database by developing a search engine in Python and querying using keywords such as "Artificial Intelligence", "Machine Learning", "Deep Learning", and "Neural Network", in addition to "Renewable Energy", and "Green Energy" within the period 2000–2021. This study focused on journals, conference papers, books, and book chapters in the English language. Scopus allows users to receive only 2000 per query. However, our developed search engine is capable of retrieving the required data without any limitation. Our developed search engine breaks the dataset to a number of chunks, each can contain up to 25 documents. In our case, there are 226 chunks containing 25 documents each and one chunk containing 11 documents. Another common method for data acquisition is employing SQL queries to search related keywords (Venugopalan and Rai, 2015; De Clercq et al., 2019). The extracted data and developed search engine is accessible on KamranNiroomand's GitHub account at: https://github.com/KamranNiroomand/Scopus-Search-Engine.git. The developed search engine can be used in the future research, using Scopus database, in various areas such as scientometrics, and intelligent decision support systems.

2.2. Pre-processing dataset

The extracted raw dataset was then prepared for the all-MiniLM-L6-v2 algorithm, a sentence-based pre-trained model. Three features were selected in this study: data, title, and abstract. Since each title has useful information about the context, the title and abstract were merged into a new feature for the analysis. Besides, the BERTopic model was employed, and unlike other topic modeling algorithm, it does not require an intensive preparation BERTopic (Grootendorst, 2022). Given the nature of BERTopic, the sentence's primary structure is necessary (Egger and Yu, 2021). Finally, the required features were converted

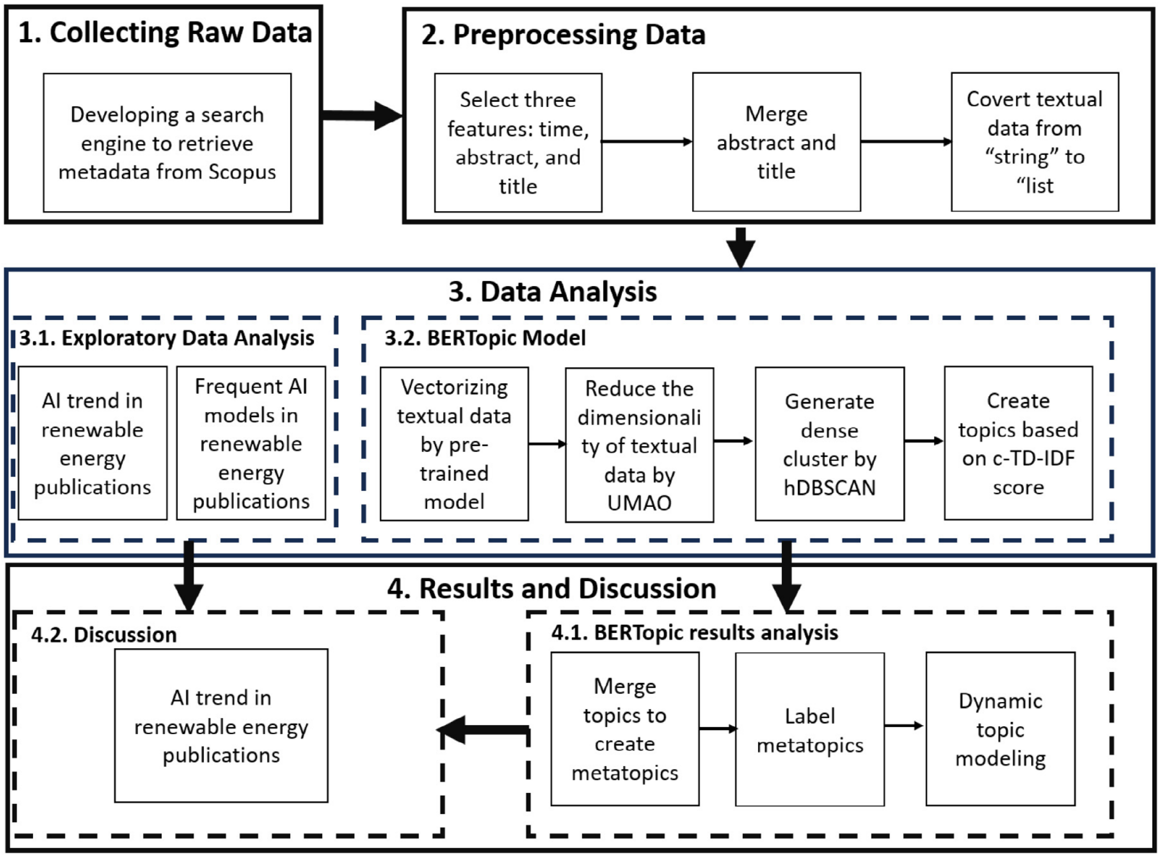

Fig. 1. Schematic of the methodology.

from a "string" type to a "list" type for the next steps. However, conventional models like LDA need comprehensive data preparation steps such as removing stop words, lemmatization, and tokenizing (Kadhim et al., 2014).

2.3. Data analysis

2.3.1. Exploratory data analysis (EDA)

Before building the BERTopic model, several data exploratory analyses were conducted, such as the rate of renewable energy systems publications in which AI has been leveraged throughout the period to investigate the trend of AI in renewable energy systems (Fig. 2). In Fig. 2, the rate is measured by dividing the number of publications that employed AI by the total number of publications in the renewable energies domain. Additionally, the most frequent modeling approaches used in renewable energy research are detected and depicted over the specified period (Fig. 3).

2.3.2. BERTopic model

This study employed BERTopic modeling algorithm (Grootendorst, 2020) to extract topics at the intersection of AI modeling and renewable energy systems and investigate them throughout 2000–2021. BERTopic is built based on Top2Vec (Angelov, 2020) and is an embedding-based model. The BERTopic model has been built in three steps. The first step vectorized the textual dataset to group close semantical terms (Egger, 2022) using a pre-trained sentence-based transformer algorithm (Reimers and Gurevych, 2019). Then, due to the high degree of sparsity within the generated vectors, a uniform various approximation and projection (UMAP) was utilized (McInnes et al., 2018) to reduce the dimensionality of vector space and keep global and local data structures. The vector space was reduced to 20 dimensions to create dense regions and employed Hierarchical Density-Based Spatial Clustering of Applications with Noise algorithm (hDBSCAN) (Campello et al., 2013; McInnes et al., 2017) as a clustering measure to identify these areas in the documents. Finally, to create topics, this study considered the c-TD-IDF algorithm, where documents in a cluster are considered one document. Then TD-IDF score is calculated to show the importance of each word in a cluster. We considered the default parameters and algorithms of the BERTopic since they gave us the best results. For example, we tried different extents of "n_neighbors" for UMAP algorithm, 15, which is the default value, 18, and 13. The result was almost the same. Also, we tried RoBERTa, a BERT-based model, instead of the all-MiniLM-L6-v2 algorithm, which is the default algorithm of BERTopic, but the results were not satisfying. Also, we set auto for "nr_topic" and True for "calculate_probabilities" and "verbose". The TD-IDF (Joachims, 1996) can be calculated using Eq. (1):

$$W_{t,d} = tf_{t,d} . \log(N/df_t) \quad (1)$$

where the term frequency, $f_{t,d}$, models the frequency of term t in document d. The inverse document frequency indicates the amount of information that a term gives to a document and is measured by the logarithm function of the number of documents in a corpus that is denoted by N divided by the total number of documents that include t. The adjusted class-based TD-IDF (Grootendorst, 2022) can be expressed as in Eq. (2):

$$W_{t,c} = tf_{t,c} . \log(1 + \mathrm{A}/tf_t) \quad (2)$$

where the term frequency, $f_{t,c}$, models the frequency of term t within a class c. The higher the value of the c-TD-IDF, the more important the words in the clusters are.

In addition, we merged topics to reduce their number from 98 to 7 dense and semantic meta-topics considering the result of the hierarchical clustering measured by the cosine distance matrix between clusters (Figure S1 supplementary information) and the domain expert's knowledge. Because of the limited capability of quantifying algorithms

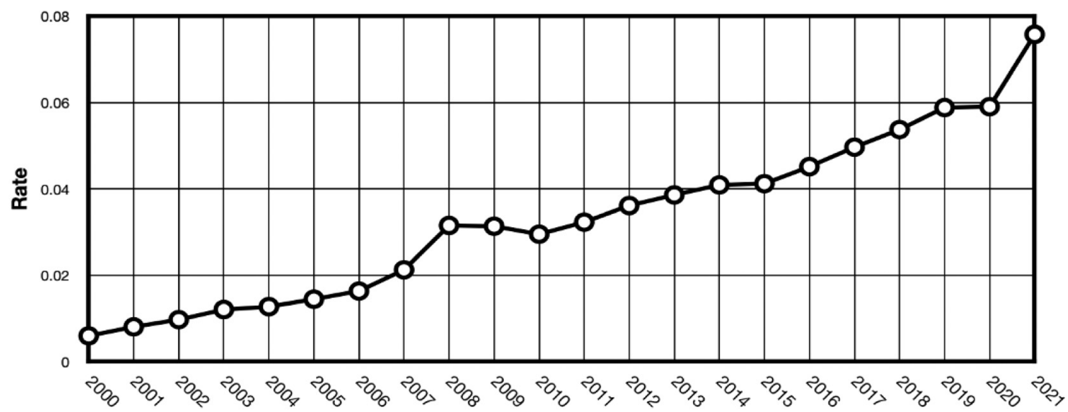

Fig. 2. Trend of AI modeling in renewable energy systems publications.

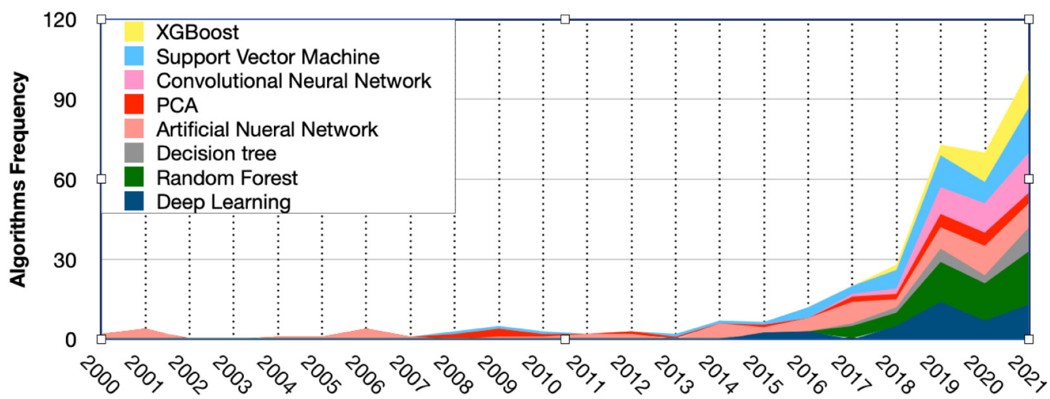

Fig. 3. Most frequent computational algorithms in renewable energy systems.

to provide sufficient contextual comprehension (Egger and Yu, 2021), topic modeling interpretation certainly needs human judgment (Hannigan et al., 2019) and domain expert knowledge (Egger and Yu, 2021). For the domain expert to choose interpretable labels for meta-topics, the top ten words of each meta-topic (Fig. 4) and the "Term Score" graph (Figure S2 supplementary information), which shows the number of words representing each topic, were taken into account. This study also includes DTM in which the identified topic fluctuation throughout the period was investigated and evolved to ascertain emerged technologies by considering unigram (one world). Blei and Lafferty (2006) first employed the DTM method, built upon LDA, to solve the static concept of topic modeling. One of the BERTopic model functions is performing DTM based on c-TD-IDF by creating global topics without considering their temporal nature (Grootendorst, 2022). To implement this, it was fitted to the whole textual dataset to create a global view of topics. Following that, local topics' representation can be generated by Eq. (3):

$$W_{t,c,i} = tf_{t,c,i} . \log(1 + \mathrm{A}/tf_t) \quad (3)$$

where the documents' term frequency is multiplied at timestep "i" considering the pre-calculated value of global IDF.

2.4. Computation

This study's programming parts were developed using the Python 3.10.2 language within the Google Colab notebook environment. It took around 25 min to run BERTopic on the merged data of "abstract" and "title" features of 5561 publications on Google Colab.

3. Results and discussion

3.1. Exploratory data analysis (EDA)

The proportion of leveraging AI toward renewable energy research is illustrated in Fig. 2. Renewable energy research has experienced an increasing trend over the considered period. It is noticeable that from 2016 the slope of the graph has increased, which shows the power of AI modeling in this sector. However, the slope became almost flat in 2019–2020, likely due to the impact of the COVID-19 pandemic and its associated restrictions (Harper et al., 2020). During that year, most of research was focused on treating methods for COVID-19 or effects of this virus on different aspects of our life (Verma and Gustafsson, 2020; Herrera Viedma, 2020). Afterward, it started to grow with a sharper slope from 2020; the proportion of renewable energy research that employed AI peaked at 0.075 of total research within this domain, nearly double its 2015 value.

This study detected and investigated the most frequent computer science methods employed in renewable energy research over the specified period. The artificial neural network has been used from the beginning of the period (Fig. 3). However, with time, researchers started to use other algorithms as well. For instance, from 2007, supervised and unsupervised methods using principal components analysis and support vector machine, respectively, emerged in renewable energy-related works. Gradually, various models appeared, such as deep learning methods in the research domain capable of handling dynamic situations like wind and wave energies with a high performance (Gu and Li, 2022). Also, deep learning algorithms, especially, deep

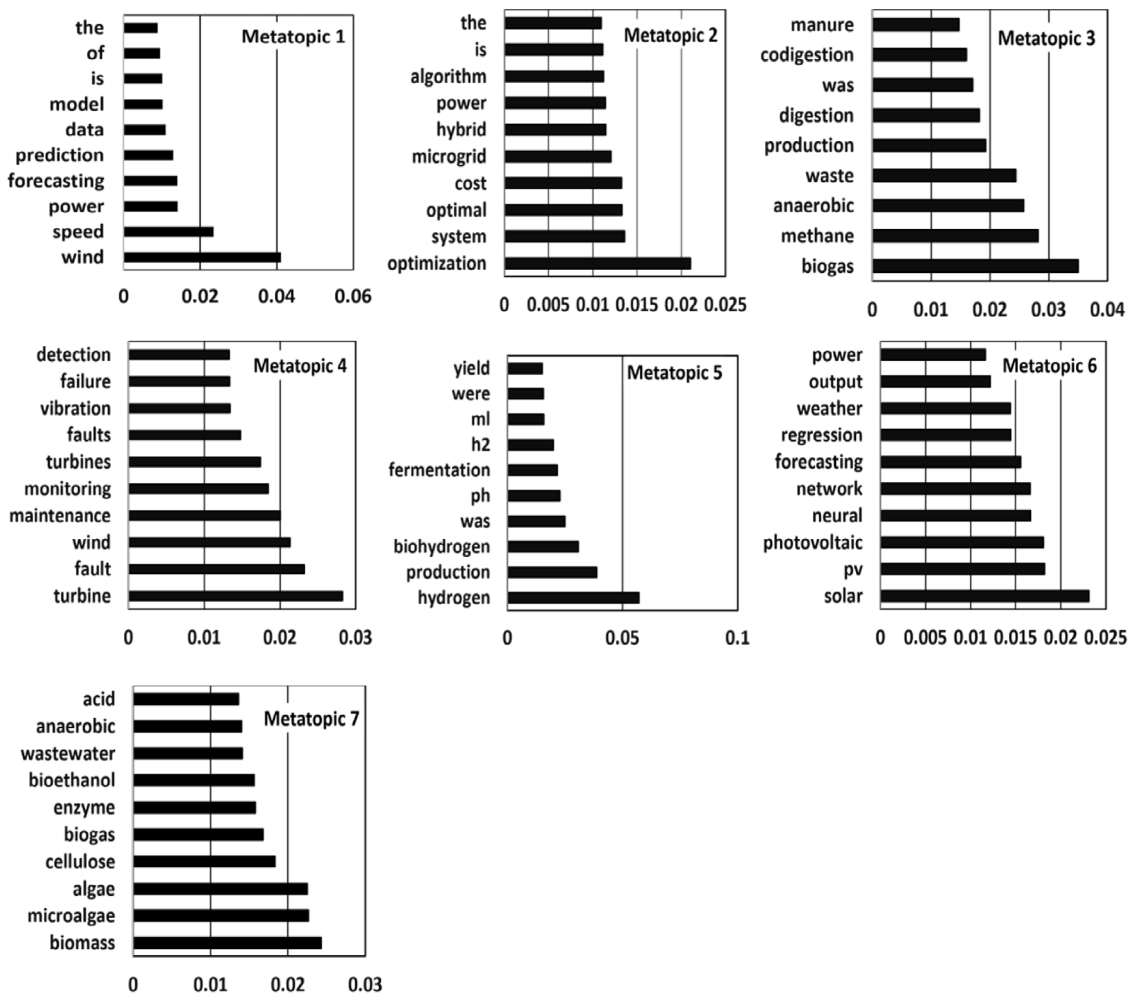

Fig. 4. c-TD-IDF for each term in seven identified meta-topics.

reinforcement learning algorithms were utilized for complex renewable energy problems such as smart grid systems with uncertainty and nonlinearity and/or large structure datasets (Widodo et al., 2021). By 2016, the decision tree emerged alongside the random forest. The latter is an ensemble learning method with several advantages over the decision tree (Ahmad et al., 2018); thus, its usage rate exceeded that of the decision tree in 2016 and increased significantly throughout the rest of the period. Besides, XGBoost, a boosting model based on a decision tree, appeared in 2017 and grew considerably from 2018 to 2021. Likewise, deep learning and convolutional neural network emerged in 2015 and 2016, respectively, and grew noticeably until the end of 2021.

3.2. Technology direction based on c-TD-IDF

Using the BERTopic model, similar topics have been merged due to clustering algorithms and a hierarchical graph. For instance, topics 12, 32, 44, 91, 58, 82, 14, 90, 10, 7, and 15 have overlaps, covering different angles of “solar energy as a renewable energy resource”. Therefore, these topics have been merged, and seven meta-topics (Prediction in Wind Systems, Power Systems Optimization, Biogas, Wind Turbine Fault Detection, Biohydrogen, Solar Energy and Photovoltaic Cells, and Biomass) have been identified. Fig. 4 shows the top 10 words corresponding to each meta-topic based on their c-TD-IDF scores. Domain experts evaluated the coherence and coverage of the generated topics based on their extensive knowledge and have chosen a human-interpretable label for each topic, considering the top-ten words of each meta-topic in terms of c-TD-IDF (Fig. 4) and term rank graph (Figure S2 supplementary information). We had three domain experts, two professors from environmental engineering and one process engineering professor, whose individual knowledge is compliment to each other. We consider odd number of domain experts to prevent possible tie of equal votes from domain experts. Their opinions on how to merge topics to create meta-topics and how to choose a name for each meta-topic were almost in line with each other. To choose the interpretable name for each meta-topic, they rewrote the most useful keywords in a meaningful way. For example, Fig. 4 shows all ten words of meta-topic 5 have a high level of relevancy to their cluster. Fig. 4 also illustrates that meta-topic 4 includes words with high scores of c-TD-IDF, such as “turbine”, “fault”, “wind”, and “detection”, implying that this meta-topic is related to wind turbine fault detection. Besides, meta-topic 3 contains “biogas”, “methane”, “anaerobic”, and “waste”, illustrating that this meta-topic should be related to leveraging AI modeling in the biogas process. Words of meta-topic 2 that have top ranks in terms of c-TD-IDF score contain “optimization”, ”microgrid”, and “algorithm”, implying power systems optimization by leveraging AI modeling techniques like deep reinforcement learning (Domínguez-Barbero et al., 2020; Ji et al., 2019). The seven identified meta-topics are (1) Prediction in Wind Systems, (2) Power Systems Optimization, (3) Biogas, (4) Wind Turbine Fault Detection, (5) Biohydrogen, (6) Solar Energy and Photovoltaic Cells, and (7) Biomass.

One of the powerful features of BERTopic is that it does not need the preprocessing of raw data, since it is an embedding-based algorithm, but this generates a large number of topics, and merging them needs knowledge and significant attention.

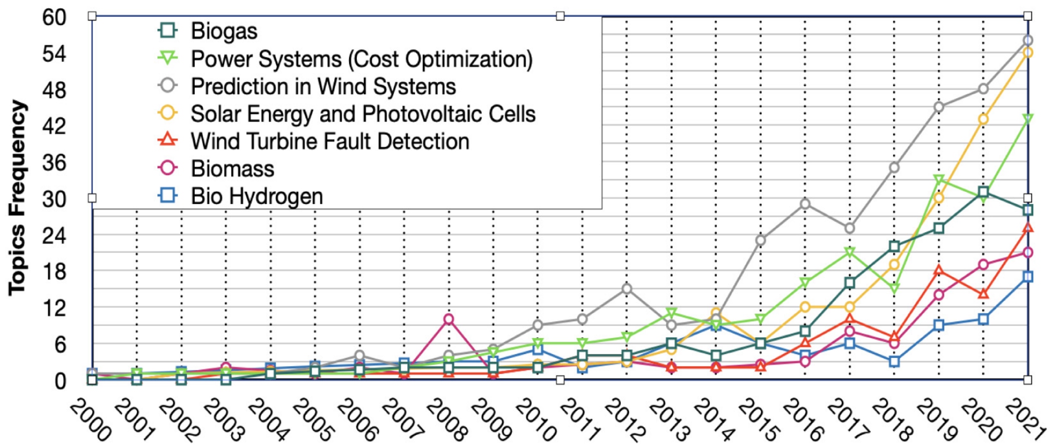

Fig. 5. Topics evolution over the period.

Fig. 5 depicts the evolution of identified meta-topics from 2000 to 2021. AI modeling emerged in renewable energy research in 2005. Afterward, leveraging computational algorithms increased at a similar rate in different types of renewable energy systems. The year 2014 was a turning point when researchers began to utilize more advanced techniques such as different deep learning algorithms. From 2014 on, the prediction in complicated systems like wind systems, solar energy and photovoltaic cells, and power systems optimization meta-topics experienced considerable growth. This aligns with our statement about implementing different deep learning algorithms, such as deep reinforcement learning, long short-term memory, and convolutional neural networks for dynamic and uncertain systems. Wind systems have been found to be a dominant and most attractive technology where deep learning models have been applied for wind speed prediction (Noorollahi et al., 2016; Moustris et al., 2016; Wang et al., 2015); (Huang et al., 2021; Yeghikian et al., 2021; Shamshirband et al., 2019) and wind power prediction (Wu et al., 2016; Zameer et al., 2017; Dong et al., 2017). State-of-the-art AI techniques play a vital role in optimizing and controlling photovoltaic and solar energy technologies (Ghannam et al., 2019) and make them more economical (Youssef et al., 2017). For instance, photovoltaic systems damages can be detected by deep convolutional neural network (Pierdicca et al., 2018), and their energy can be forecasted by a novel deep learning architecture (Abdel-Basset et al., 2021).

Similarly, deep learning has various applications in power systems, such as online energy scheduling (Ji et al., 2021), power systems resilience improvement (Kamruzzaman et al., 2021), iterative learning control optimization (Zhou et al., 2022), hybrid-driven fuzzy filtering optimization (Zhang et al., 2023), and adaptive power system emergency control (Huang et al., 2020). Biogas production is a waste-based technology through anaerobic digestion mainly for generating renewable energy and also for the valorization of organic residues (Kougias and Angelidaki, 2018; Appels et al., 2011). Chen et al. (2017) mentioned that the optimal design of biogas plants depends on local conditions, namely, substrate supply and local infrastructure. Hence, developing machine learning/deep learning algorithms that can determine optimal conditions for industrial-scale plants, the value of each feedstock (Chiu et al., 2022), and predicting the outcome of biogas production is necessary from an economic point of view. In other words, improving biogas plant economic sustainability will end up in decreasing operational costs. Considering the increasingly importance of industrial biogas facilities, the significant growth of available data in this, and the capability of AI-based techniques in enhancing the economic sustainability of biogas facilities and, therefore, this area will probably emerge as one of the main prevalent domains of renewable energy systems. Biohydrogen is one the most environmentally friendly fuels since its combustion produces H_2O as a carbon-free by-product and is also generated under anaerobic conditions and without consuming fossil fuels (Brentner et al., 2010). However, this process has several challenges, such as the cost of hydrogen as fuel, infrastructure facilities, distribution, and storage (Kamaraj et al., 2019). Despite optimizing AI modeling in biohydrogen systems (Wang et al., 2021; Khaleghi et al., 2021; Lian et al., 2021; Liu et al., 2021), these challenges showed their impact on the prevalence of biohydrogen topic. Continuous research and collaboration of data scientists and engineers would result in improving biohydrogen technology and utilizing it in hydrogen-based vehicles (Manoharan et al., 2019).

Regarding biomass, alongside forecasting biomass characteristics, process outcome, and performance of bio-energy end-use systems, one of the primary usages of AI is to generate synthetic datasets (Liao and Yao, 2021). Data augmentation will increase the amount of labeled data and enhance the performance of supervised machine learning algorithms (Bowles et al., 2018). The generative datasets in biomass are particularly useful in terms of biomass properties, biofuel properties, kinetic parameters, engine performance, and Life Cycle Inventory (LCI) (Liao and Yao, 2021). Considering the fast pace of AI development as well as growing sustainable systems data scientists, it can be predicted that analytic measures and AI integrated models would significantly overcome the challenges such as lack of sufficient high-quality data in the biomass domain, which leads to comprehensive assessment and optimization of biomass systems. The AI-based feature of this study that extracts topics automatically, diminishes subjectivity from the exercise and enables consistent and comprehensive comparisons between topics and between time intervals. Seven extracted meta-topics provide comprehensive and logical coverage of the research field and have a high level of similarity to the topics being used to produce review paper studies (Lateef et al., 2022) or scientometric research (Sohail et al., 2022) in renewable energy domains.

DTM is a probabilistic time-series-based model capable of analyzing the evolution of topics over timeframes (Blei and Lafferty, 2006). DTM has various applications in different disciplines. For instance, Ayele and Juell-Skielse (2020) investigated the evolution of self-driving cars' topics and trends from 2000 to 2019. They chose DTM since, unlike LDA, DTM considers the temporal aspect of each topic (how topics have evolved over the period). Their results illustrated the evolution of twenty topics in the self-driving car domain, including software system architecture and design, brake system and safety, and navigation in self-driving. Lee et al. (2016) applied dynamic topic modeling to toxicogenomics data. It was used as an alternative technique to discover underlying patterns in time-series gene expression profiles which results in gaining a perception of the dynamic behavior of genes in the related systems. Besides, Morimoto and Kawasaki (2017) utilized dynamic topic modeling as a forecasting method for financial market volatility, which enhanced forecasting accuracy. Linton et al. (2017)

used dynamic topic modeling into cryptocurrency community forums to investigate the evolution of different topics related to big events in the cryptocurrency society. Tabassum et al. (2021) used dynamic topic modeling to analyze social media by focusing on hashtags. Guldi (2019) applied a dynamic topic modeling algorithm to build history record of British infrastructure.

Because the gap between scientific literature and commercialization is narrowing, text mining of academic papers can play a significant role for both stakeholders intending to invest in renewable energy sectors as well as data-driven start-ups working on renewable energy systems. Notably, the most realistic technology road mapping with a broad scope can be achieved by validating the findings of text mining with the results of existing scientific publications and patent documents under the supervision of domain experts.

This current study is limited to scientific publications from the Scopus database from 2000 to 2021 to characterize the landscape of AI in renewable energy systems research. Future studies could consider other databases, such as the Web of Science, PubMed, IEEE Xplore, ScienceDirect or other data sources, such as patents, to perform complementary exploration. In addition, in the current study, we only analyzed uni-grams. Future studies can also consider the whole body of the paper for their analysis; particularly, the methodology section should contain informative text regarding methodological evolution. Additionally, we only considered English documents and future studies can expand their scope by analyzing published documents in other languages and using other NLP techniques such as GPT3 and WuDao 2.0 for their analyses.

Summary

This study targeted and implemented a prototype pipeline for automating data extraction and clustering. To do so, we developed a capable search engine based on Scopus API, which is able to break the limit of the Scopus database and can retrieve unlimited documentation per query. In addition, we investigated AI in renewable energy systems publications throughout the 21's century from two points of view. First, by analyzing term frequency, we investigated the trend of most frequent computational algorithms in renewable energy systems papers. The result showed that, within recent years, researchers started leveraging more diverse and complicated methods like deep learning techniques. However, still, conventional models including random forest are more popular. We also uncovered the latent research topics of AI in renewable energy systems and considered their temporal aspect by employing the DTM method. Our DTM's results demonstrated the role of AI-based models in overcoming uncertainty and enhancing risk management in solar and wind systems Tawn and Browell (2022). The generated results reveal particular attention to modeling and simulation research projects. Matching and comparing the results of both employed computational models and DTM analyses demonstrate that recent advancements in computational science have built new pathways for complicated renewable energy system problems. Some examples can be energy storage optimization of wind, solar, and photovoltaic systems (Abualigah et al., 2022), and monitoring and anomaly detection of wind turbines (Xiang et al., 2022).

This feature that most steps of the proposed pipeline are automatic is significant. In other words, insights can be produced quickly, even for complex research fields involving a large number of papers annually, to assist policymakers and decision-makers in understanding the research dynamics better and help with research and development (R&D) strategies. This is of particular importance for R&D that needs high pace progress (Ebadi et al., 2022), for instance, disruptive technology development that can affect strategic stability (Sechser et al., 2019), national security (Ebadi et al., 2022), and economic development (Rifkin, 2011). Therefore, organizations that can better understand and monitor the research landscape will have a competitive edge.

4. Conclusions

The research characterized and mapped AI applications in renewable energy systems by leveraging a text mining technique to build semantic and dense structure clusters in a semantically continuous space. The reason behind conducting this study is establishing a technology roadmap and research perspective by building a reliable and efficient pipeline that extracts hidden patterns and provides decision-makers with new and useful insights. In this regard, the study used an unstructured dataset of 5661 scientific works between 2000 and 2021. This work contributes comprehensively to identifying technologies at the intersection of AI and renewable energy systems and enhancing previous scientific works in pattern detection by leveraging novel algorithms for the NLP and the BERTopic method, resulting in a high level of coherency and efficiency. More specifically, it saves the time computationally because BERTopic does not require time-consuming parts such as intensive data preprocessing and hyperparameter tuning to analyze textual datasets. The study investigated technology trends by considering c-TD-IDF, annual analyses of topic evolution, most frequent AI algorithms, and the number of renewable energy systems papers that employed AI. BERTopic showed seven dense meta-topics covering all aspects of various types of renewable energy systems. The c-TD-IDF score and term score decline graphs provide insightful information to discriminate meta-topics and label them interpretably. For instance, meta-topics 7 is associated with terms such as biomass, microalgae, algae, and cellulose, which gained the highest c-TD-IDF. Additionally, the term score decline graph proves the previous statement as the first four terms of this meta-topics are reliable to be considered for labeling. The analysis showed that the development of AI modeling significantly impacted areas associated with a high level of uncertainty, such as wind and microgrid systems. Future studies can expand their scope by considering other databases (e.g., US patent, ScienceDirect, and Google Scholar) or their combinations and considering published documents in other languages. Also, they can analyze bi-gram and tri-gram instead of considering only uni-gram. Besides, instead of focusing only on the abstract and title, they can consider the entire body of the document.

CRediT authorship contribution statement

Kamran Niroomand: Conceptualization, Methodology, Writing – original draft, Software, Data curation, Visualization, Investigation. **Noori M. Cata Saady:** Conceptualization, Writing – original draft, Investigation, Supervision. **Carlos Bazan:** Editing, Supervision, Validation. **Sohrab Zendehboudi:** Writing – original draft, Validation. **Amilcar Soares:** Methodology, Software, Investigation, Writing – review & editing. **Talib M. Albayati:** Editing, Writing – review & editing.

Declaration of competing interest

The authors declare no financial interests/personal relationships which may be considered as potential competing interests.

Data availability

The data is shared on Github and a link is provided in the manuscript.

Acknowledgments

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number 528169-2019], and NSERC Discovery Grant for anaerobic digestion of nitrogen-rich feedstock, Canada, and the Government of Newfoundland, Canada and Labrador through the Canadian Agriculture Partnership administrated by the Department of Fisheries and Land Resources, Canada. The funding sources have no involvement in the particulars of the article.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.engappai.2023.106848.

References

Abdel-Basset, M., Hawash, H., Chakrabortty, R.K., Ryan, M., 2021. PV-Net: An innovative deep learning approach for efficient forecasting of short-term photovoltaic energy production. J. Clean. Prod. 303, 127037. http://dx.doi.org/10.1016/j.jclepro.2021.127037.

Abualigah, L., Zitar, R.A., Almotairi, K.H., Hussein, A.M., Abd Elaziz, M., Nikoo, M.R., Gandomi, A.H., 2022. Wind, solar, and photovoltaic renewable energy systems with and without energy storage optimization: a survey of advanced machine learning and deep learning techniques. Energies 15 (2), 578.

Ahmad, M.W., Mourshed, M., Rezgui, Y., 2018. Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. Energy 164, 465–474. http://dx.doi.org/10.1016/j.energy.2018.08.207.

Amer, M., Daim, T.U., 2010. Application of technology roadmaps for renewable energy sector. Technol. Forecast. Soc. Change 77 (8), 1355–1370. http://dx.doi.org/10.1016/j.techfore.2010.05.002.

Angelo, A.C.M., Saraiva, A.B., Clímaco, J.C.N., Infante, C.E., Valle, R., 2017. Life cycle assessment and multi-criteria decision analysis: Selection of a strategy for domestic food waste management in Rio de Janeiro. J. Clean. Prod. 143, 744–756. http://dx.doi.org/10.1016/j.jclepro.2016.12.049.

Angelov, D., 2020. Top2Vec: Distributed representations of topics. Retrieved from http://arxiv.org/pdf/2008.09470v1.

Appels, L., Assche, A.V., Willems, K., Degrève, J., Impe, J.V., Dewil, R., 2011. Peracetic acid oxidation as an alternative pre-treatment for the anaerobic digestion of waste activated sludge. Bioresour. Technol. 102 (5), 4124–4130. http://dx.doi.org/10.1016/j.biortech.2010.12.070.

Ayele, W.Y., Juell-Skielse, G., 2020. Eliciting evolving topics, trends and foresight about self-driving cars using dynamic topic modeling. In: Future of Information and Communication Conference. Springer, Cham, pp. 488–509.

Azarpour, A., Mohammadzadeh, O., Rezaei, N., Zendehboudi, S., 2022. Current status and future prospects of renewable and sustainable energy in North America: Progress and challenges. Energy Convers. Manage. 269, 115945.

Beltramo, T., Klocke, M., Hitzmann, B., 2019. Prediction of the biogas production using GA and ACO input features selection method for ANN model. Inf. Process. Agric. 6 (3), 349–356. http://dx.doi.org/10.1016/j.inpa.2019.01.002.

Blei, D.M., Lafferty, J.D., 2006. Dynamic topic models. In: Proceedings of the 23rd International Conference on Machine Learning - ICML '06. The 23rd International Conference. ACM Press, http://dx.doi.org/10.1145/1143844.1143859.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3 (Jan), 993–1022.

Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R., Hammers, A., Dickie, D.A., Hernández, M.V., Wardlaw, J., Rueckert, D., 2018. GAN augmentation: Augmenting training data using generative adversarial networks (version 1). http://dx.doi.org/10.48550/ARXIV.1810.10863, arXiv.

Brentner, L.B., Peccia, J., Zimmerman, J.B., 2010. Challenges in developing biohydrogen as a sustainable energy source: Implications for a research Agenda. Environ. Sci. Technol. 44 (7), 2243–2254. http://dx.doi.org/10.1021/es9030613.

Campello, R.J.G.B., Moulavi, D., Sander, J., 2013. Density-based clustering based on hierarchical density estimates. In: Advances in Knowledge Discovery and Data Mining. Springer Berlin Heidelberg, pp. 160–172. http://dx.doi.org/10.1007/978-3-642-37456-2_14.

Capizzi, G., Lo Sciuto, G., Napoli, C., Woźniak, M., Susi, G., 2020. A spiking neural network-based long-term prediction system for biogas production. Neural Netw. 129, 271–279. http://dx.doi.org/10.1016/j.neunet.2020.06.001.

Chen, L., Cong, R.-G., Shu, B., Mi, Z.-F., 2017. A sustainable biogas model in China: The case study of Beijing deqingyuan biogas project. Renew. Sustain. Energy Rev. 78, 773–779. http://dx.doi.org/10.1016/j.rser.2017.05.027.

Chen, X., He, T., Hu, X., An, Y., Wu, X., 2011. Inferring functional groups from microbial gene catalogue with probabilistic topic models. In: IEEE International Conference on Bioinformatics and Biomedicine. BIBM, pp. 3–9.

Chiu, M.-C., Wen, C.-Y., Hsu, H.-W., Wang, W.-C., 2022. Key wastes selection and prediction improvement for biogas production through hybrid machine learning methods. Sustain. Energy Technol. Assess. 52, 102223. http://dx.doi.org/10.1016/j.seta.2022.102223.

Chowdhary, K.R., 2020. Natural language processing. Fundam. Artif. Intell. 603–649. http://dx.doi.org/10.1007/978-81-322-3972-7_19.

Coupé, T., 2003. Science is golden: Academic R&D and university patents. J. Technol. Transf. 28 (1), 31–46. http://dx.doi.org/10.1023/a:1021626702728.

Daenekindt, S., Huisman, J., 2020. Mapping the scattered field of research on higher education. a correlated topic model of 17, 000 articles, 1991–2018. Higher Educ. 80 (3), 571–587. http://dx.doi.org/10.1007/s10734-020-00500-x.

De Clercq, D., Wen, Z., Song, Q., 2019. Innovation hotspots in food waste treatment, biogas, and anaerobic digestion technology: a natural language processing approach. Sci. Total Environ. 673, 402–413. http://dx.doi.org/10.1016/j.scitotenv.2019.04.051.

Domínguez-Barbero, D., García-González, J., Sanz-Bobi, M.A., Sánchez-Úbeda, E.F., 2020. Optimising a microgrid system by deep reinforcement learning techniques. Energies 13 (11), 2830. http://dx.doi.org/10.3390/en13112830.

Dong, Q., Sun, Y., Li, P., 2017. A novel forecasting model based on a hybrid processing strategy and an optimized local linear fuzzy neural network to make wind power forecasting: A case study of wind farms in China. Renew. Energy 102, 241–257. http://dx.doi.org/10.1016/j.renene.2016.10.030.

Dong, B., Xu, G., Luo, X., Cai, Y., Gao, W., 2012. A bibliometric analysis of solar power research from 1991 to 2010. Scientometrics 93 (3), 1101–1117. http://dx.doi.org/10.1007/s11192-012-0730-9.

Ebadi, A., Auger, A., Gauthier, Y., 2022. On the evolution of research in hypersonics: application of natural language processing and machine learning (version 1). http://dx.doi.org/10.48550/ARXIV.2208.08507, arXiv.

Egger, R. (Ed.), 2022. Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications. Springer Nature, http://dx.doi.org/10.1007/978-3-030-88389-8_16.

Egger, R., Yu, J., 2021. Identifying hidden semantic structures in instagram data: a topic modelling comparison. Tour. Rev. http://dx.doi.org/10.1108/tr-05-2021-0244.

Gallagher, R.J., Reing, K., Kale, D., Ver Steeg, G., 2017. Anchored correlation explanation: Topic modeling with minimal domain knowledge. Trans. Assoc. Comput. Linguist. 5, 529–542. http://dx.doi.org/10.1162/tacl_a_00078.

Ghannam, R., Klaine, P.V., Imran, M., 2019. Artificial intelligence for photovoltaic systems. Power Syst. 121–142. http://dx.doi.org/10.1007/978-981-13-6151-7_6.

Gonçalves Neto, J., Vidal Ozorio, L., Campos de Abreu, T.C., Ferreira dos Santos, B., Pradelle, F., 2021. Modeling of biogas production from food, fruits and vegetables wastes using artificial neural network (ANN). Fuel 285, 119081. http://dx.doi.org/10.1016/j.fuel.2020.119081.

Grootendorst, M., 2020. Topic modeling with BERT. | Towards data science. Retrieved from https://towardsdatascience.com/topic-modeling-with-bert-779f7db187e6.

Grootendorst, M., 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure (version 1). http://dx.doi.org/10.48550/ARXIV.2203.05794, arXiv.

Gu, C., Li, H., 2022. Review on deep learning research and applications in wind and wave energy. Energies 15 (4), 1510. http://dx.doi.org/10.3390/en15041510.

Guldi, J., 2019. Parliament's debates about infrastructure: An exercise in using dynamic topic models to synthesize historical change. Technol. Cult. 60 (1), 1–33. http://dx.doi.org/10.1353/tech.2019.0000.

Hannigan, T.R., Haans, R.F.J., Vakili, K., Tchalian, H., Glaser, V.L., Wang, M.S., Kaplan, S., Jennings, P.D., 2019. Topic modeling in management research: Rendering new theory from textual data. Acad. Manag. Ann. 13 (2), 586–632. http://dx.doi.org/10.5465/annals.2017.0099.

Hansen, B.D., Tamouk, J., Tidmarsh, C.A., Johansen, R., Moeslund, T.B., Jensen, D.G., 2020. Prediction of the methane production in biogas plants using a combined gompertz and machine learning model. In: Computational Science and Its Applications. ICCSA 2020, Springer International Publishing, pp. 734–745. http://dx.doi.org/10.1007/978-3-030-58799-4_53.

Harper, L., Kalfa, N., Beckers, G.M.A., Kaefer, M., Nieuwhof-Leppink, A.J., Fossum, M., Herbst, K.W., Bagli, D., 2020. The impact of COVID-19 on research. J. Pediatr. Urol. 16 (5), 715–716. http://dx.doi.org/10.1016/j.jpurol.2020.07.002.

Herrera Viedma, E., 2020. Global trends in coronavirus research at the time of Covid-19: A general bibliometric approach and content analysis using scimat.

Huang, Q., Huang, R., Hao, W., Tan, J., Fan, R., Huang, Z., 2020. Adaptive power system emergency control using deep reinforcement learning. IEEE Trans. Smart Grid 11 (2), 1171–1182. http://dx.doi.org/10.1109/tsg.2019.2933191.

Huang, L., Kang, J., Wan, M., Fang, L., Zhang, C., Zeng, Z., 2021. Solar radiation prediction using different machine learning algorithms and implications for extreme climate events. Front. Earth Sci. 9, 596860.

Jallan, Y., Brogan, E., Ashuri, B., Clevenger, C.M., 2019. Application of natural language processing and text mining to identify patterns in construction-defect litigation cases. J. Leg. Aff. Disput. Resolut. Eng. Constr. 11 (4), 04519024. http://dx.doi.org/10.1061/(asce)la.1943-4170.0000308.

Ji, Y., Wang, J., Xu, J., Fang, X., Zhang, H., 2019. Real-time energy management of a microgrid using deep reinforcement learning. Energies 12 (12), 2291. http://dx.doi.org/10.3390/en12122291.

Ji, Y., Wang, J., Xu, J., Li, D., 2021. Data-driven online energy scheduling of a microgrid based on deep reinforcement learning. Energies 14 (8), 2120. http://dx.doi.org/10.3390/en14082120.

Jiang, H., Qiang, M., Lin, P., 2016. A topic modeling based bibliometric exploration of hydropower research. Renew. Sustain. Energy Rev. 57, 226–237. http://dx.doi.org/10.1016/j.rser.2015.12.194.

Joachims, T., 1996. A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization. Carnegie-mellon univ pittsburgh pa dept of computer science.

Jockers, M.L., Thalken, R., 2020. Topic modeling. In: Text Analysis with R. Springer International Publishing, pp. 211–235. http://dx.doi.org/10.1007/978-3-030-39643-5_17.

Johri, A., Olds, B.M., 2011. Situated engineering learning: Bridging engineering education research and the learning sciences. J. Eng. Educ. 100 (1), 151–185.

Jusoh, S., 2018. A study on NLP applications and ambiguity problems. J. Theor. Appl. Inf. Technol. 96 (6).

Kadhim, A.I., Cheah, Y.-N., Ahamed, N.H., 2014. Text document preprocessing and dimension reduction techniques for text document clustering. In: 2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology. 2014 Artificial Intelligence with Applications in Engineering and Technology. ICAIET, IEEE, http://dx.doi.org/10.1109/icaiet.2014.21.

Kamaraj, M., Ramachandran, K.K., Aravind, J., 2019. Biohydrogen production from waste materials: benefits and challenges. Int. J. Environ. Sci. Technol. 17 (1), 559–576. http://dx.doi.org/10.1007/s13762-019-02577-z.

Kamruzzaman, Md, Duan, J., Shi, D., Benidris, M., 2021. A deep reinforcement learning-based multi-agent framework to enhance power system resilience using shunt resources. IEEE Trans. Power Syst. 36 (6), 5525–5536. http://dx.doi.org/10.1109/tpwrs.2021.3078446.

Khaleghi, M.K., Savizi, I.S.P., Lewis, N.E., Shojaosadati, S.A., 2021. Synergisms of machine learning and constraint-based modeling of metabolism for analysis and optimization of fermentation parameters. Biotechnol. J. 16 (11), 2100212. http://dx.doi.org/10.1002/biot.202100212.

Kougias, P.G., Angelidaki, I., 2018. Biogas and its opportunities—A review. Front. Environ. Sci. Eng. 12 (3), http://dx.doi.org/10.1007/s11783-018-1037-8.

Kyebambe, M.N., Cheng, G., Huang, Y., He, C., Zhang, Z., 2017. Forecasting emerging technologies: A supervised learning approach through patent analysis. Technol. Forecast. Soc. Change 125, 236–244. http://dx.doi.org/10.1016/j.techfore.2017.08.002.

Landauer, T.K., Foltz, P.W., Laham, D., 1998. An introduction to latent semantic analysis. Discourse Process. 25 (2–3), 259–284. http://dx.doi.org/10.1080/01638539809545028.

Lateef, A.A.A., Ali Al-Janabi, S.I., Abdulteef, O.A., 2022. Artificial intelligence techniques applied on renewable energy systems: A review. In: Proceedings of International Conference on Computing and Communication Networks. Springer Nature Singapore, pp. 297–308. http://dx.doi.org/10.1007/978-981-19-0604-6_25.

Lee, M., Liu, Z., Huang, R., Tong, W., 2016. Application of dynamic topic models to toxicogenomics data. BMC Bioinformatics 17 (S13), http://dx.doi.org/10.1186/s12859-016-1225-0.

Lee, J., Yi, J.-S., Son, J., 2019. Development of automatic-extraction model of poisonous clauses in international construction contracts using rule-based NLP. J. Comput. Civ. Eng. 33 (3), 04019003. http://dx.doi.org/10.1061/(asce)cp.1943-5487.0000807.

Li, G., Liu, M., Zhang, X., Wang, C., Lai, K.H., Qian, W., 2022. Semantic recognition of ship motion patterns entering and leaving port based on topic model. J. Mar. Sci. Eng. 10 (12), 2012.

Lian, Z., Wang, Y., Zhang, X., Yusuf, A., Famiyeh, Lord, Murindababisha, D., Jin, H., Liu, Y., He, J., Wang, Y., Yang, G., Sun, Y., 2021. Hydrogen production by fluidized bed reactors: A quantitative perspective using the supervised machine learning approach. J 4 (3), 266–287. http://dx.doi.org/10.3390/j4030022.

Liao, M., Yao, Y., 2021. Applications of artificial intelligence-based modeling for bioenergy systems: A review. GCB Bioenerg. 13 (5), 774–802. http://dx.doi.org/10.1111/gcbb.12816.

Lindstedt, N.C., 2019. Structural topic modeling for social scientists: A brief case study with social movement studies literature, 2005–2017. Soc. Curr. 6 (4), 307–318. http://dx.doi.org/10.1177/2329496519846505.

Linton, M., Teo, E.G.S., Bommes, E., Chen, C.Y., Härdle, W.K., 2017. Dynamic topic modelling for cryptocurrency community forums. Appl. Quant. Financ. 355–372. http://dx.doi.org/10.1007/978-3-662-54486-0_18.

Liu, Y., Liu, J., He, H., Yang, S., Wang, Y., Hu, J., Jin, H., Cui, T., Yang, G., Sun, Y., 2021. A review of enhancement of biohydrogen productions by chemical addition using a supervised machine learning method. Energies 14 (18), 5916. http://dx.doi.org/10.3390/en14185916.

Mahmoodi-Eshkaftaki, M., Ebrahimi, R., 2021. Integrated deep learning neural network and desirability analysis in biogas plants: A powerful tool to optimize biogas purification. Energy 231, 121073. http://dx.doi.org/10.1016/j.energy.2021.121073.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., et al., 2018. Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. Commun. Methods Meas. 12 (2–3), 93–118.

Manoharan, Y., Hosseini, S.E., Butler, B., Alzhahrani, H., Senior, B.T.F., Ashuri, T., Krohn, J., 2019. Hydrogen fuel cell vehicles; current status and future prospect. Appl. Sci. 9 (11), 2296. http://dx.doi.org/10.3390/app9112296.

McInnes, L., Healy, J., Astels, S., 2017. Hdbscan: Hierarchical density based clustering. J. Open Source Softw. 2 (11), 205. http://dx.doi.org/10.21105/joss.00205.

McInnes, L., Healy, J., Melville, J., 2018. UMAP: Uniform manifold approximation and projection for dimension reduction (version 3). http://dx.doi.org/10.48550/ARXIV.1802.03426, arXiv.

Morimoto, T., Kawasaki, Y., 2017. Forecasting financial market volatility using a dynamic topic model. Asia-Pac. Financ. Mark. 24 (3), 149–167. http://dx.doi.org/10.1007/s10690-017-9228-z.

Mosallaie, S., Rad, M., Schiffauerova, A., Ebadi, A., 2021. Discovering the evolution of artificial intelligence in cancer research using dynamic topic modeling. COLLNET J. Scientometr. Inf. Manag. 15 (2), 225–240. http://dx.doi.org/10.1080/09737766.2021.1958659.

Moustris, K.P., Zafirakis, D., Alamo, D.H., Nebot Medina, R.J., Kaldellis, J.K., 2016. 24-H ahead wind speed prediction for the optimum operation of hybrid power stations with the use of artificial neural networks. Perspect. Atmospheric Sci. 409–414. http://dx.doi.org/10.1007/978-3-319-35095-0_58.

Najafi, B., Faizollahzadeh Ardabili, S., 2018. Application of ANFIS, ANN, and logistic methods in estimating biogas production from spent mushroom compost (SMC). Resour. Conserv. Recy. 133, 169–178. http://dx.doi.org/10.1016/j.resconrec.2018.02.025.

Noorollahi, Y., Jokar, M.A., Kalhor, A., 2016. Using artificial neural networks for temporal and spatial wind speed forecasting in Iran. Energy Convers. Manage. 115, 17–25. http://dx.doi.org/10.1016/j.enconman.2016.02.041.

Pierdicca, R., Malinverni, E.S., Piccinini, F., Paolanti, M., Felicetti, A., Zingaretti, P., 2018. Deep convolutional neural network for automatic detection of damaged photovoltaic cells. Int. Arch. Photogramm. Remote. Sens. Spatial Inf. Sci. 42 (2).

Poege, F., Harhoff, D., Gaessler, F., Baruffaldi, S., 2019. Science quality and the value of inventions. Sci. Adv. 5 (12), http://dx.doi.org/10.1126/sciadv.aay7323.

Reimers, N., Gurevych, I., 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. Available online at: https://arxiv.org/abs/1908.10084.

Reisenbichler, M., Reutterer, T., 2019. Topic modeling in marketing: recent advances and research opportunities. J. Bus. Econ. 89 (3), 327–356.

Ren, Y., Yu, M., Wu, C., Wang, Q., Gao, M., Huang, Q., Liu, Y., 2018. A comprehensive review on food waste anaerobic digestion: Research updates and tendencies. Bioresour. Technol. 247, 1069–1076. http://dx.doi.org/10.1016/j.biortech.2017.09.109.

Rifkin, J., 2011. The Third Industrial Revolution: How Lateral Power Is Transforming Energy, the Economy, and the World. Macmillan.

Sechser, T.S., Narang, N., Talmadge, C., 2019. Emerging technologies and strategic stability in peacetime, crisis, and war. J. Strateg. Stud. 42 (6), 727–735.

Seo, K.W., Seo, J., Kim, K., Ji Lim, S., Chung, J., 2021. Prediction of biogas production rate from dry anaerobic digestion of food waste: Process-based approach vs. recurrent neural network black-box model. Bioresour. Technol. 341, 125829. http://dx.doi.org/10.1016/j.biortech.2021.125829.

Seyyedattar, M., Ghiasi, M.M., Zendehboudi, S., Butt, S., 2020. Determination of bubble point pressure and oil formation volume factor: Extra trees compared with LSSVM-CSA hybrid and ANFIS models. Fuel 269, 116834.

Shamshirband, S., Rabczuk, T., Chau, K.W., 2019. A survey of deep learning techniques: application in wind and solar energy resources. IEEE Access 7, 164650-164666.

Sohail, M., Afrouzi, H.N., Mehranzamir, K., Ahmed, J., Siddique, M.B.M., Tabassum, M., 2022. A comprehensive scientometric analysis on hybrid renewable energy systems in developing regions of the world. Results Eng. 16, 100481.

Souili, A., Cavallucci, D., Rousselot, F., 2015. A lexico-syntactic pattern matching method to extract idm- triz knowledge from on-line patent databases. Procedia Eng. 131, 418–425. http://dx.doi.org/10.1016/j.proeng.2015.12.437.

Sun, L., Yin, Y., 2017. Discovering themes and trends in transportation research using topic modeling. Transp. Res. C 77, 49–66. http://dx.doi.org/10.1016/j.trc.2017.01.013.

Tabassum, S., Gama, J., Azevedo, P., Teixeira, L., Martins, C., Martins, A., 2021. Dynamic topic modeling using social network analytics. Prog. Artif. Intell. 498–509. http://dx.doi.org/10.1007/978-3-030-86230-5_39.

Tawn, R., Browell, J., 2022. A review of very short-term wind and solar power forecasting. Renew. Sustain. Energy Rev. 153, 111758. http://dx.doi.org/10.1016/j.rser.2021.111758.

Tazikeh, S., Zendehboudi, S., Ghafoori, S., Lohi, A., Mahinpey, N., 2022. Algal bioenergy production and utilization: Technologies, challenges, and prospects. J. Environ. Chem. Eng. 10 (3), 107863.

Tran, B.X., Latkin, C.A., Sharafeldin, N., Nguyen, K., Vu, G.T., Tam, W.W.S., Cheung, N.-M., Nguyen, H.L.T., Ho, C.S.H., Ho, R.C.M., 2019. Characterizing artificial intelligence applications in cancer research: A latent Dirichlet allocation analysis. JMIR Med. Inf. 7 (4), e14401. http://dx.doi.org/10.2196/14401.

Tufaner, F., Demirci, Y., 2020. Prediction of biogas production rate from anaerobic hybrid reactor by artificial neural network and nonlinear regressions models. Clean Technol. Environ. Policy 22 (3), 713–724. http://dx.doi.org/10.1007/s10098-020-01816-z.

Venugopalan, S., Rai, V., 2015. Topic based classification and pattern identification in patents. Technol. Forecast. Soc. Change 94, 236–250. http://dx.doi.org/10.1016/j.techfore.2014.10.006.

Verma, S., Gustafsson, A., 2020. Investigating the emerging COVID-19 research trends in the field of business and management: A bibliometric analysis approach. J. Bus. Res. 118, 253–261.

Wang, L., Long, F., Liao, W., Liu, H., 2020. Prediction of anaerobic digestion performance and identification of critical operational parameters using machine learning algorithms. Bioresour. Technol. 298, 122495. http://dx.doi.org/10.1016/j.biortech.2019.122495.

Wang, J., Qin, S., Zhou, Q., Jiang, H., 2015. Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China. Renew. Energy 76, 91–101. http://dx.doi.org/10.1016/j.renene.2014.11.011.

Wang, Y., Tang, M., Ling, J., Wang, Y., Liu, Y., Jin, H., He, J., Sun, Y., 2021. Modeling biohydrogen production using different data driven approaches. Int. J. Hydrogen Energy 46 (58), 29822–29833. http://dx.doi.org/10.1016/j.ijhydene.2021.06.122.

Widodo, D.A., Iksan, N., Udayanti, E.D., 2021. Renewable energy power generation forecasting using deep learning method. IOP Conf. Ser. Earth Environ. Sci. 700 (1), 012026.

Wu, W., Chen, K., Qiao, Y., Lu, Z., 2016. Probabilistic short-term wind power forecasting based on deep neural networks. In: 2016 International Conference on Probabilistic Methods Applied to Power Systems. PMAPS, IEEE, http://dx.doi.org/10.1109/pmaps.2016.7764155.

Xiang, L., Yang, X., Hu, A., Su, H., Wang, P., 2022. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks. Appl. Energy 305, 117925.

Xu, W., Liu, X., Gong, Y., 2003. Document clustering based on non-negative matrix factorization. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval - SIGIR '03. The 26th Annual International ACM SIGIR Conference. ACM Press, http://dx.doi.org/10.1145/860435.860485.

Yang, H.L., Chang, T.W., Choi, Y., 2018. Exploring the research trend of smart factory with topic modeling. Sustainability 10 (8), 2779.

Yang, Y., Zheng, S., Ai, Z., Jafari, M.M.M., 2021. On the prediction of biogas production from vegetables, fruits, and food wastes by ANFIS- and LSSVM-based models. In: Baghban, A. (Ed.), BioMed Research International, Vol. 2021. Hindawi Limited, pp. 1–8. http://dx.doi.org/10.1155/2021/9202127.

Yeghikian, M., Ahmadi, A., Dashti, R., Esmaeilion, F., Mahmoudan, A., Hoseinzadeh, S., Garcia, D.A., 2021. Wind farm layout optimization with different hub heights in Manjil wind farm using particle swarm optimization. Appl. Sci. 11 (20), 9746.

Youssef, A., El-Telbany, M., Zekry, A., 2017. The role of artificial intelligence in photovoltaic systems design and control: A review. Renew. Sustain. Energy Rev. 78, 72–79. http://dx.doi.org/10.1016/j.rser.2017.04.046.

Zameer, A., Arshad, J., Khan, A., Raja, M.A.Z., 2017. Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks. Energy Convers. Manage. 134, 361–372. http://dx.doi.org/10.1016/j.enconman.2016.12.032.

Zareei, S., Khodaei, J., 2017. Modeling and optimization of biogas production from cow manure and maize straw using an adaptive neuro-fuzzy inference system. Renew. Energy 114, 423–427. http://dx.doi.org/10.1016/j.renene.2017.07.050.

Zendehboudi, S., Rezaei, N., Lohi, A., 2018. Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review. Appl. Energy 228, 2539–2566.

Zhang, Z., Song, X., Sun, X., Stojanovic, V., 2023. Hybrid-driven-based fuzzy secure filtering for nonlinear parabolic partial differential equation systems with cyber attacks. Internat. J. Adapt. Control Signal Process. 37 (2), 380–398.

Zhou, C., Tao, H., Chen, Y., Stojanovic, V., Paszke, W., 2022. Robust point-to-point iterative learning control for constrained systems: A minimum energy approach. Internat. J. Robust Nonlinear Control 32 (18), 10139–10161.