

Article

A Topic Modeling Approach to Discover the Global and Local Subjects in Membrane Distillation Separation Process

Ersin Aytac¹  and Mohamed Khayet^{2,3,*} 

¹ Department of Environmental Engineering, Zonguldak Bülent Ecevit University, 67100 Zonguldak, Türkiye; ersin.aytac@beun.edu.tr

² Department of Structure of Matter, Thermal Physics and Electronics, Faculty of Physics, University Complutense of Madrid, Avda. Complutense s/n, 28040 Madrid, Spain

³ Madrid Institute for Advanced Studies of Water (IMDEA Water Institute), Calle Punto Net N° 4, 28805 Alcalá de Henares, Madrid, Spain

* Correspondence: khayetm@fis.ucm.es

Abstract: Membrane distillation (MD) is proposed as an environmentally friendly technology of emerging interest able to aid in the resolution of the worldwide water issue and brine processing by producing distilled water and treating high-saline solutions up to their saturation with a view toward reaching zero liquid discharge (ZLD) at relatively low temperature requirements and a low operating hydrostatic pressure. Topic modeling (TM), which is a Machine Learning (ML) method combined with Natural Language Processing (NLP), is a customizable approach that is ideal for researching massive datasets with unknown themes. In this study, we used BERTopic, a new cutting-edge Python library for topic modeling, to explore the global and local themes in the MD separation literature. By using the BERTopic model, the words describing the collected dataset were detected together with over- and underexplored research topics to guide MD researchers in planning their future works. The results indicated that two global themes are widely discussed and are relevant to MD scientists abroad. In brief, these topics are permeate flux, heat-energy recovery, surface modification, and polyvinylidene fluoride hydrophobic membranes. BERTopic discovered 62 local concepts. The most researched local topics were solar applications, membrane scaling, and electrospun membranes, while the least investigated were boron removal, dairy effluent applications, and nickel wastewater treatment. In addition, the topics were illustrated in a 2D plane to better understand the obtained results.

Keywords: membrane distillation; water treatment; machine learning; maximal marginal relevance; all-mpnet-base-v2; c-TF-IDF; cosine similarity; CountVectorizer; dimensionality reduction; document embedding; HDBSCAN; UMAP



Citation: Aytac, E.; Khayet, M. A Topic Modeling Approach to Discover the Global and Local Subjects in Membrane Distillation Separation Process. *Separations* **2023**, *10*, 482. <https://doi.org/10.3390/separations10090482>

Academic Editor: Victoria Samanidou

Received: 5 August 2023

Revised: 24 August 2023

Accepted: 30 August 2023

Published: 2 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The demand for fresh water continues to increase due to the rapid growth in the human population and other issues related to accelerated industrialization, environmental impacts, climate change, altered consumption patterns, etc., which lead to water stress and scarcity. Over the last century, the global water demand has expanded by 600%, and this rate equates to an annual increase of 1.8%. As a result, there is an urgent need for effective water management and conservation practices as well as for the development of developing creative environmentally friendly methods and solutions to the said global water crisis [1–3]. Membrane distillation (MD), a non-isothermal separation technology, has been proposed for clean water production and treatment of high-salinity waters up to their saturation, thus allowing the management of brines discharged from other water-processing plants and seeking to achieve zero liquid discharge, which represents an important environmental benefit. The driving force in MD is the water vapor pressure difference established at both sides of a hydrophobic porous membrane [4,5]. Among the various advantages of the MD separation process, one can highlight its potential to overcome the osmotic pressure

limitation of the aqueous solutions to be treated (i.e., this osmotic pressure limits the treatment of aqueous solutions with other processes such as the pressure-driven membrane separation process reverse osmosis (RO)), low temperature requirements (i.e., MD can be applied at temperatures below the boiling point of the aqueous feed solutions, thus low-temperature solar energy systems and industrial waste heat can be used), and low operating pressures (i.e., MD can be used at atmospheric pressure), among others [6–9]. In addition, MD separation can reach a nearly 100% non-volatile solute rejection rate [10]. Furthermore, different types of hydrophobic membranes (polyvinylidene fluoride (PVDF), polytetrafluoroethylene (PTFE), or polypropylene (PP); single-layer or multi-layered flat sheet, hollow fiber, or nanofibrous membranes; etc.) were employed [11]. In general, other than membrane engineering, this technology has been studied by researchers following different theoretical and experimental aspects, like the effects of all involved operating parameters on the MD performance, the application of different MD configurations (i.e., direct contact MD (DCMD), air gap MD (AGMD), sweeping gas MD (SGMD), vacuum MD (VMD), etc.), the treatment of different water sources such as sea and brackish waters and industrial wastewaters (food, pharmaceutical, radioactive, municipal, dyes, acids, heavy metals, oils, etc.), the development of different theoretical models and simulations (e.g., computational fluid dynamics (CFD), Artificial Neural Networks (ANNs), numerical analysis, etc.) and Machine Learning (ML), among others [12–19].

Artificial intelligence (AI) is the computer science sub-branch that can provide computers and machines with cognitive abilities to learn, draw conclusions, and make decisions based on a collection of data [20,21]. The transformational power of AI may be found in a variety of industries with different applications [22]. Although classically it is very difficult to define the coverage of AI, Natural Language Processing (NLP), Text Mining (TM), Robotics, Machine Learning (ML), Rule-based or Knowledge-Based Systems, Case-Based Reasoning, Neural Networks (NNs), and Computer Game Playing are located under its heading. ML, a sub-branch of AI, is the process of extracting meaningful information from data and encompasses computers that can be trained to help people with little or no sustained effort [23]. The last decade has witnessed an increase and success in the use of ML in various fields; for instance, forecasting the future, making movie suggestions, recommending products to buy, deciding on loan applications, affecting hiring decisions, etc. [24,25]. ML algorithms are broadly classified into three types: supervised learning, unsupervised learning, and reinforcement learning [26]. Supervised methods like classification or regression can train a classifier to predict new information using a defined input and output. Unsupervised learning techniques like clustering and dimension reduction (DR) can uncover hidden patterns in data. Reinforcement ML can learn from past experiences and find the best actions to take in an unfamiliar environment to achieve the ideal state transition for achieving a specific goal [27].

Topic modeling (TM), a combination of Natural Language Processing (NLP) and Machine Learning (ML), is generally used to identify and extract the main concepts that are discussed in a collection of text documents. TM can find hidden patterns (i.e., that cannot be identified directly by humans) and forms in large collections of unstructured text data. Topic modeling is also very adaptable, as it can be applied to various types of text data. The number of topics, for example, can be changed to make the themes more or less specific or accurate, and the model can be fine-tuned to better suit specific characteristics of the data. It has received a lot of attention since it was first introduced by Blei in 2003 [28–30]. When it is employed as an unsupervised learning method, it does not require labeled data or prior knowledge of the topics in the data. This makes it useful for exploring and comprehending large datasets when the concepts are initially unknown [31,32]. Unsupervised learning methods are data-driven approaches that reveal internal data characteristics and laws through the learning of unlabeled data sets. These approaches excavate the internal features of the data in greater depth, making it more conducive to extracting discriminative features [33,34].

Many methods have been developed to find existent latent topics in a given dataset. Non-Negative Matrix Factorization (NMF), Latent Semantic Analysis (LSA), probabilistic LSA, and Latent Dirichlet Allocation (LDA) are among these techniques [35]. In 2022, a new state-of-the-art model—BERTopic—was introduced by Maarten Grootendorst [36] for topic modeling as a Python library. This generalized model for pretrained sentence transformers has yielded promising results for topic modeling in a variety of domains [37]. The BERTopic model assigns only one topic per document and aids in the identification of outlier documents that are difficult to classify, resulting in an improved classification accuracy [38]. The BERTopic model is based on six phases (the first five steps are mandatory, while the last step is optional). A related illustration can be seen in Figure 1 [39].

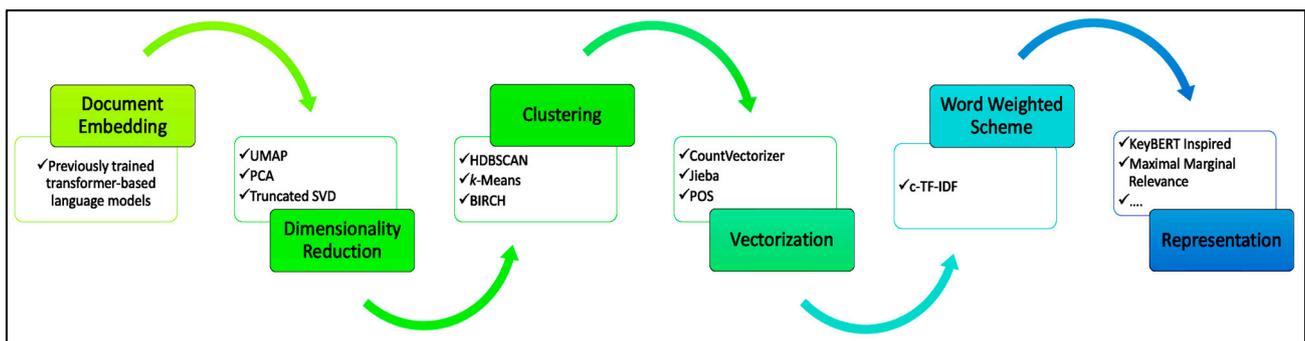


Figure 1. Basic process sequence of the BERTopic algorithm used.

As can be seen in Figure 1, the first step of the procedure is document embedding, which is the core of the sequence in a text-based intelligent system. It is the process of converting a textual input into a numerical array (vector) form for applying ML models, and it uses structure-preserving maps to capture informative representations from high-dimensional observations [40–42]. BERTopic makes use of any transformer-based language models that have been previously trained. The following second step is a Dimensionality Reduction (DR) approach [37]. DR techniques are applied to reduce the number of input features in a set of data, which becomes more compact, improving the efficiency of the learning algorithm [43]. DR can help users reduce the data storage space, decrease the computational time of ML algorithms, and help visualize multidimensional data in lower dimensions such as 2D or 3D [44,45]. Many unsupervised Dimensionality Reduction methods, such as Multiple Dimensional Scaling (MDS), Principal Component Analysis (PCA), Locally Linear Embedding (LLE), Isometric Mapping (ISOMAP), and Uniform Manifold Approximation and Projection (UMAP), have been proposed in the literature [46]. The third step includes the algorithm clusters of the reduced embeddings [47]. Clustering in ML is a dynamic technique of categorizing data into numerous collections or clusters based on the similarities of the data points’ characteristics and features [48,49]. Conventional clustering techniques are known as “unsupervised”, indicating that no information about data point partitioning or outcome variables is available [50]. Clustering approaches are categorized into two types: hierarchical and partitional. Hierarchical clustering attempts to form a tree-like layout of classes and partition occurrences in each node of the tree, whereas partitioning clustering categorizes occurrences effectively into k clusters [51]. The algorithms used for clustering can be k -means, spectral, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), and Ordering Points to Identify the Clustering Structure (OPTICS). These algorithms can detect the underlying structures in image, text, or video [48,52]. Subsequently, in step 4, the BERTopic model will tokenize and vectorize documents. The quality of topic representations is critical in topic modeling for interpreting topics, communicating results, and understanding patterns. It is critical to ensure that the topic representations are appropriate for a given case [39]. Tokenization is the process of breaking down text strings into small chunks such as words and phrases [53]. Because

ML models only accept matrices as inputs, the unstructured data must be converted to vectors. The technique of translating text into numerical form is known as “text vectorization”. In this case, Term Frequency—Inverse Document Frequency (TF-IDF), Doc2Vec, and CountVectorizers are commonly used vectorizers for textual data [54,55]. The fifth step is necessary to obtain an accurate representation of the discovered themes and to reflect the important words in the clusters. A Combined Term Frequency–Inverse Document Frequency (c-TF-IDF) method is used. c-TF-IDF, a modified version of TF-IDF, considers what distinguishes documents in one cluster from documents in another. Finally, in the last step, the representation of the topics is the fine-tuning. This optional step allows users to represent the concepts with more unique keywords. The BERTopic architecture offers a wide spectrum of options for fine-tuning that ranges from KeyBERT-like models to GPT-like models [39].

MD research started in 1967 when the first paper, “Vaporization through Porous Membrane”, was published by Findley, M. E. [56]. It has a widespread 56-year history with thousands of manuscripts written by many worldwide researchers. This non-isothermal separation process has been discussed from numerous perspectives, such as laboratory experiments, the production of innovative membranes, system improvement, theoretical modeling, and optimizations, among others. In our previous study, trending topics in MD literature were identified via bibliometric methods, Text Mining approaches, and manual searching [14], but none of the published papers revealed the research concepts in MD with a recently developed, state-of-the-art AI model. This paper employed the BERTopic algorithm to discover the most attractive and interesting MD research subjects based on the provided abstracts of the articles downloaded from the Scopus database. Several insights about MD can be revealed using the topic modeling approach: (i) the predominantly handled and the less discussed topics by MD researchers; (ii) not only the information on MD topics can be provided, but also information regarding the prominent or lagging MD configurations, membrane types, and modeling approaches, among others, on any given topic (thanks to the list of topic terms created to enable in-depth exploration of the topic terms; and (iii) a topic modeling of MD literature can reveal new perspectives or insights that might not have been noticed before by identifying the gaps. One can easily identify the relationships between different techniques and applications by looking to the BERTopic results. In addition, this kind of approach to MD helps to guide scientists to carry out pioneering and cutting-edge MD research studies. As a result, topic modeling is a valuable and effective tool for improving knowledge and innovation in MD. This study also includes a detailed description and application of the BERTopic procedure, which we believe will inspire further studies on MD or in any other scientific domain.

2. Data and Methods

2.1. Data

In this research, we used an MD dataset downloaded 23 January 2023 from the Scopus database. This database was chosen because it is well known in the academic community for its broad, inclusive, and comprehensive content coverage [57,58]. The search criteria and the keywords can be seen in Tables 1 and 2, respectively.

Table 1. Search criteria followed to download the MD dataset.

Criteria	Description
Title–Abstract–Keyword	<i>Limit to</i> Keyword List (Table 2)
Source Type	<i>Limit to</i> Journal
Document Type	<i>Limit to</i> Article
Publication Stage	<i>Limit to</i> Final
Language	<i>Limit to</i> English
Publication Year	<i>Exclude</i> 2023

Table 2. Keywords considered in collecting the MD dataset.

Keyword	Keyword
membrane distillation	MD
air gap membrane distillation	AGMD
direct contact membrane distillation	DCMD
vacuum membrane distillation	VMD
vacuum enhanced membrane distillation	VEMD
Sweeping/sweep gas membrane distillation	SGMD
membrane air stripping	MAS
thermostatic sweeping gas membrane distillation	TSGMD
permeate gap membrane distillation	PGMD
liquid gap membrane distillation	LGMD
water gap membrane distillation	WGMD
conductive gap membrane distillation	CGMD
material gap membrane distillation	MGMD

In addition, those articles that were not found in the search results were manually added to the dataset. The collection was then manually screened to remove irrelevant documents and articles that did not have abstracts. The final dataset included 3684 articles.

2.2. Methods

The BERTopic architecture was used to reveal the hidden themes in the MD domain. BERTopic used 5 consecutive ML approaches to uncover the topics in the collection. The first approach was to transform the textual data into numerical representations (text embedding). The BERTopic architecture has a structure that allows many different pretrained embedding models. In this study, the selected embedding model was *all-mpnet-base-v2* because it was the highest-quality model at the time of the research (i.e., the highest performance on sentence embedding for 14 different datasets with the highest average performance). *all-mpnet-base-v2* is an all-around model optimized for a wide range of applications. Over 1 billion training pairs were used to train this model on a huge and diverse dataset. *all-mpnet-base-v2* uses a mean pooling approach with normalized embeddings, and it converts a data instance into a 768-dimension (feature) numerical array [59].

As stated earlier, after the conversion of the dataset to numerical representations, the BERTopic library applied a DR technique. For this step, UMAP was used. This is a non-linear method that reduces dimensionality using manifold learning and topological data analysis. When reducing dimensionality, UMAP preserves the data’s local and global structure, which is critical for capturing the semantics of textual data. The data is compressed to low dimensions (mostly 2D or 3D) by attempting to minimize the cross-entropy (CE). The basic calculation of CE is as follows (Equation (1)) [60]:

$$CE = \sum_{a \in A} \mu(a) \log \left(\frac{\mu(a)}{v(a)} \right) + (1 - \mu(a)) \log \left(\frac{1 - \mu(a)}{1 - v(a)} \right) \tag{1}$$

where A is the weighted adjacency matrix derivation of z , and μ and v represent two types of probabilities. Here, z ($z = \{z_1, z_2, \dots, z_n\}, z_n \in \mathbb{R}^M$) is the lower-dimensional representation of the high-dimensional dataset x . The details of the UMAP calculations can be found elsewhere (McInnes et al. [61]).

The data were then fed to a clustering algorithm for segmentation. The HDBSCAN method was used for the data clustering. HDBSCAN is an updated version of DBSCAN with varying epsilon values (ϵ) that integrates the results to identify the optimal clustering for stability across the epsilon. This enables HDBSCAN to detect clusters of various densities and to be more resilient in parameter selection. The HDBSCAN clustering algorithm exhibits important advantages over other clustering algorithms, as it produces a separate cluster for outliers, reduces the amount of noise in the clusters, and determines the number of clusters automatically [62]. The computational path of HDBSCAN starts

with the two hyperparameters that the algorithm needs for clustering: ϵ , the distance scale; and k , the minimum number of points. $X = \{X_1, \dots, X_n\}$ is a set of points in a metric space with the Euclidian distance d and $X_i \in X$. X_i , the core point of cluster i within ϵ , is at least equal to k , as indicated in Equation (2) [63,64]:

$$|B(X_i, \epsilon) \cap X| \geq k \tag{2}$$

where B is the open ball radius.

In addition to X_i and X_j , the two arbitrary points are considered as ϵ – reachable, depending on ϵ and k (Equations (3) and (4)):

$$X_i \in B(X_j, \epsilon) \tag{3}$$

$$X_j \in B(X_i, \epsilon) \tag{4}$$

When all the data instances in a cluster are connected, a cluster is formed. The HDBSCAN algorithm uses a modified distance metric. The core-distance ($\kappa(X_i)$) is the distance of X_i to its k^{th} nearest neighbors, and the equation that describes the mutual reachability distance between X_i and X_j ($d_{mreach}(X_i, X_j)$) is as follows:

$$d_{mreach}(X_i, X_j) = \begin{cases} \max\{\kappa(X_i), \kappa(X_j), d(X_i, X_j)\} & X_i \neq X_j \\ 0 & X_i = X_j \end{cases} \tag{5}$$

The detected outliers are moved further away from clusters by the mutual reachability distance. By applying the traditional Single Linkage Clustering algorithm to the discrete metric space, the hierarchical clustering of X is established. A clustering on density variation can be used to discover regions with the highest density inside a point cloud, where the local density at each point is calculated by estimating the core-distance value associated with each point (Equation (6)):

$$\lambda = \frac{1}{\epsilon} \tag{6}$$

The cluster tree’s hierarchy can be reduced by recursively merging some of the clusters. The cluster tree is condensed by taking the minimum permissible cluster size (m) and only admitting the pruning of a cluster that would not endure against the increment in the λ value into at least two subsets with sizes greater than m . According to this method, the stability (σ) of an individual cluster is defined by adding the range of λ values for each point in the cluster, as written in Equation (7):

$$\sigma(C_i) = \sum_{X_j \in C_i} (\lambda_{max,C_i}(X_j) - \lambda_{min,C_i}(X_j)) \tag{7}$$

In this case, $\lambda_{max,C_i}(X_j)$ and $\lambda_{min,C_i}(X_j)$ are the bounds of λ over the point X_j in the cluster C_i . To obtain the best clustering attribution among all conceivable clustering results, the overall persistence score in all selected clusters should be maximized while considering the constraint of no cluster overlap. Clusters with the highest total persistency were chosen for this purpose as indicated by the following equation:

$$\sum_{i \in I} \sigma(C_i) \tag{8}$$

where I is the subset of the total number of clusters (n). For all $i, j \in I$ and $i \neq j$, Equation (8) is limited by Equation (9):

$$C_i \cap C_j = \emptyset \tag{9}$$

In the next step, the data were tokenized and vectorized. BERTopic uses CountVectorizer as the default for these purposes. First, each textual data instance is tokenized, and then the text is converted into features (F). If there are (D) documents and (F) features, CountVectorizer will convert them into a $D \times F$ matrix. The values in the matrix represent the frequency of each feature [65].

In the fifth step, the BERTopic method allowed the representation of topics of the clusters based on the c-TF-IDF approach. Equation (1) was used to calculate the c-TF-IDF ($W_{x,c}$) of a single word. For the term x within class c , the c-TF-IDF score can be calculated using Equation (10) [36]:

$$W_{x,c} = \left\| tf_{x,c} \right\| x \log \left(1 + \frac{A}{f_x} \right) \tag{10}$$

where $tf_{x,c}$ is the frequency of the word x in a class c , f_x is the frequency of the word x across all classes, and A is the average number of words per class.

In the last step, which is optional, the researcher can apply a technique called Maximal Marginal Relevance (MMR) to reduce the word repetition and increase keyword diversity (fine-tuning of topic representations). MMR, a good approach to present non-redundant information, considers the similarity of key words within the document as well as the similarity of previously picked phrases. MMR is defined as follows (Equation (11)) [66]:

$$MMR = Arg \max_{D_i \in R/S} \left[\lambda Sim_1(D_i, F) - (1 - \lambda) \max_{d_j \in S} Sim_2(D_i, D_j) \right] \tag{11}$$

where D represents the sentence collection, F is the feature set, and S is the subset of sentences in D that have already been chosen. In D , R/S is the set of unselected sentences and λ is the diversification constant, which is a float (0–1). Sim_1 measures the similarity between a sentence and a feature, whereas Sim_2 measures the similarity between two phrases.

Apart from the 6 steps mentioned above, BERTopic can create different plots to interpret the obtained results. The most important of these plots is the heatmap of the topic’s similarity. The heatmap created by the model is based on cosine similarity (CS). The basic idea behind cosine similarity is to compute the cosine value of the angle between two vectors to demonstrate their similarity. CS ranges between -1 and 1 . The cosine similarity value is equal to 1 when two vectors point in the same direction and -1 when they point in opposite directions. For two vectors ($X^V = \{x_1, \dots, x_n\}$, $Y^V = \{y_1, \dots, y_n\}$), the CS can be computed as follows (Equation (12)) [67]:

$$CS(X, Y) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \tag{12}$$

3. Results and Discussion

3.1. Outline of the MD Dataset

Before diving in the MD topic modeling results, it is critical to provide key information about the dataset, since it helps to comprehend the nature and quality. MD separation is a promising non-isothermal technology in the field of desalination and water treatment. It is a subject of increasing interest worldwide for numerous research groups. The annual publication count can provide an indication of the degree of research activity in a specific collection and can be used to track its growth and advancement over time. The evidence of this rising attention is shown in Figure 2.

As shown in Figure 2, the number of MD publications has increased exponentially, particularly since 2012. A total of 473 articles were published in 2022, the last year of the dataset. The total number of articles (3684) indicates that membranologists are making active efforts to produce appropriate MD membranes with improved performance and high thermal efficiency to optimize MD systems. However, as the number of publications increases, the question regarding which topics are being covered more and which are being

ignored becomes more important. For this reason, it is extremely informative to determine those MD topics of great interest for future studies using the BERTopic architecture.

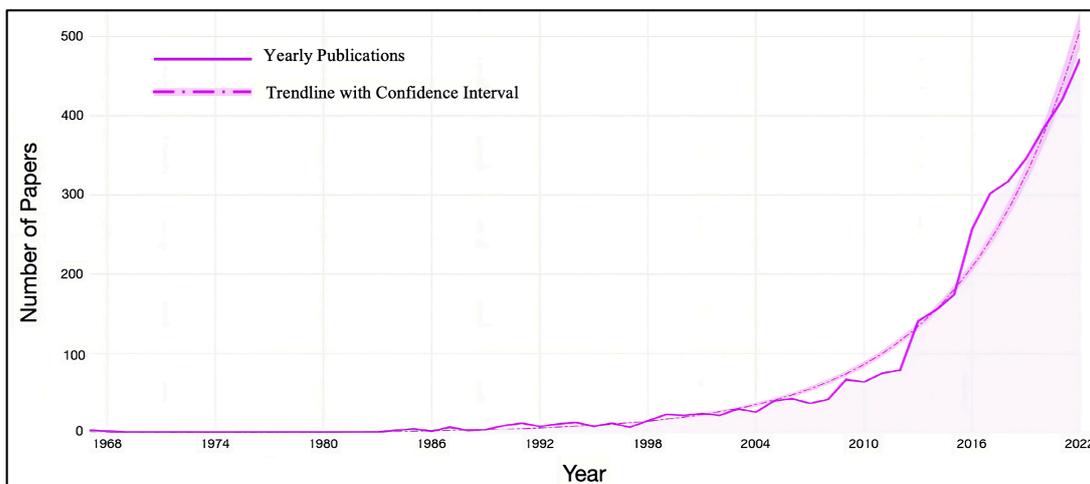


Figure 2. Yearly MD article publications.

The distribution of some significant metrics of the domain was revealed. The advantage of using a violin plot to visualize the distribution of numerical values is that it can help to summarize the basic statistics as well as to show the density of each variable. Figure 3 illustrates the violin plots of the MD publication year, citation, page count, and reference count values of the domain; the main statistical values (min, max, mean, median, outliers, first quartile, and third quartile) are presented. The wide areas in the figures reflect more frequent data points, while the thin sections represent the less frequent data points.

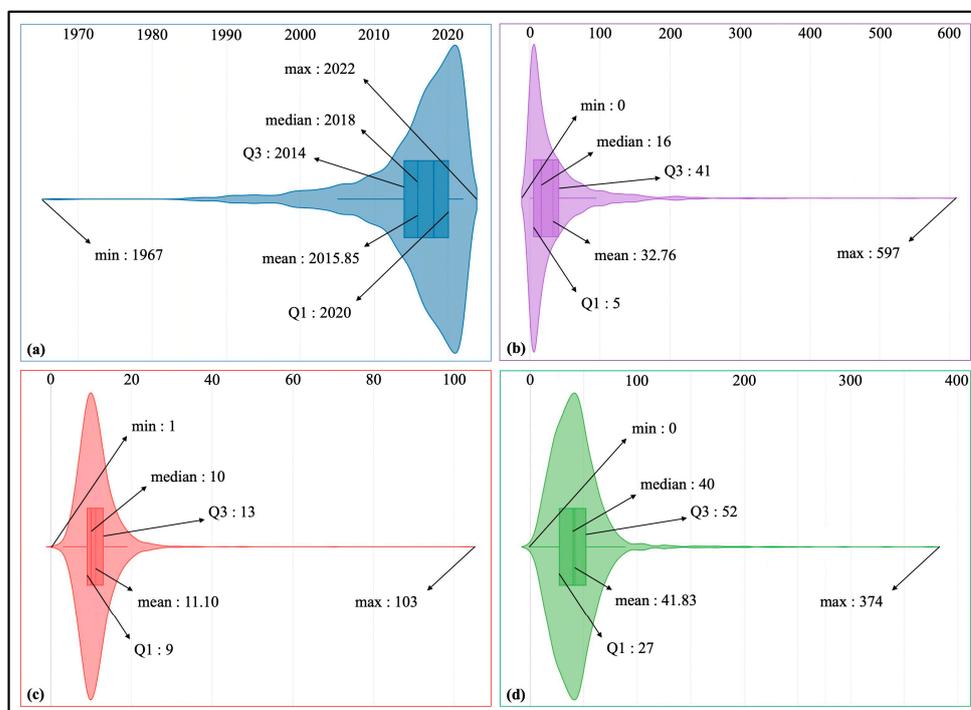


Figure 3. Violin plots of significant MD values: (a) publication year; (b) times cited; (c) page count; (d) reference count.

Figure 3a contains information about the years of publication of the MD articles in the dataset. The oldest published article on MD was dated 1967 [56]. However, since then MD did not gain much attention in the scientific community until 2012, when an exponential

increase in MD studies was initiated. The mean (2015.85) and median (2018) values of the publication years are also evidence of this interest. When the dataset was divided in half using the median value, the sum of the studies conducted before 2018 was equal to the number of studies conducted during the last 4 years. When the number of citations received by MD articles was analyzed (Figure 3b), it could be seen that an article received ~33 citations on average, which is reasonably good. There are articles in the collection that had not been cited at all (i.e., zero citations), but it should be noted that this was to be expected for those articles published in the last months of 2022. It is clear that Figure 3b is left-skewed (which is the same as right-tailed), indicating that there were articles with more than 150 citations, and these publications were seen as outliers (i.e., very impactful papers with more citations than expected) in the dataset. Figure 3c shows the violin plot of the number of pages and page distributions of the articles in the MD domain. In Figure 3c, it can be seen that the articles with more than 30 pages protrude to the right in the graph. While most of the publications contained between 8–12 pages, the average value was ~11. Finally, when the violin plot of the reference count was examined (Figure 3d), the data were better distributed than the other features when looking at the width of the green region. While the average reference value was ~42, the median value (40) indicated that the number of articles using more than 40 references was equal to the number of articles using less than 40 references.

3.2. Terms Defining the MD Domain

In the BERTopic model, analysts can find the themes that define global topics and indicate specialized themes (i.e., local topics). The most important parameter in the BERTopic architecture to create this variance is *min_cluster_size* (in HDBSCAN algorithm), which is the primary parameter that affects the resulted clustering. Ideally, this is a simple option to configure the lowest-sized grouping (i.e., the number of clusters that will be generated) for which researchers want to consider a cluster. Increasing this parameter results in fewer but larger clusters, whilst reducing this value results in more microclusters [68,69].

To find the words that describes the MD domain, the *min_cluster_size* parameter was set to 3684, which naturally resulted in one cluster. The indicated cluster was described with the following words (Table 3) that also defined the collection. Furthermore, the dataset was illustrated in a two-dimensional space (Figure 4).

Table 3. The resulting top 10 terms that defined the collection (*min_cluster_size* = 3684).

Rank	Term
1	membrane
2	water
3	distillation
4	flux
5	membranes
6	MD
7	feed
8	process
9	temperature
10	performance

Table 3 shows the terms mostly used by MD community (i.e., words describing the published research studies) and the related improvement efforts to make this technology a worldwide separation process. The term “dimension” in this figure depicts a feature that describes an aspect of the data. Dimensions 1 and 2 represent the most important features of 768 dimensions created in the first step of the algorithm (text embedding) with the *all-mpnet-base-v2* model and then reduced to two in the second step with the UMAP algorithm. In fact, researchers involved in MD are already familiar with these terms. A typical MD system consists of a high-temperature feed channel and a low-temperature permeate channel in which the vapor flux is driven by a temperature difference across

a hydrophobic and microporous membrane. The vapor flux is collected in the permeate side of the membrane. The permeate flux is an essential indicator of a given MD system's performance, since it reveals how efficiently the system produces distilled water [70,71]. In addition to the very high rejection factor, MD researchers pay special attention to the permeate flux enhancement. In a typical MD configuration, there is the feed side where the aqueous solution to be treated is heated; then the permeate (i.e., distillate) is collected either at the permeate side of the membrane inside the membrane module (e.g., DCMD), on a condensation cold plate (e.g., AGMD), or outside the membrane module (e.g., VMD, SGMD); while the non-volatile components are retained by the membrane [72]. MD research studies are mostly conducted for desalination of seawater or brackish water, although other aqueous feed solutions such as pharmaceutical, radioactive, etc., have also been considered [14]. The flow rate, chemical composition, temperature, and other properties of feed solutions are the most studied parameters in MD because of their important effects on MD performance [73,74]. In fact, the temperature of both the feed and permeate solutions are the main parameters affecting MD performance, especially the MD permeate flux and the thermal efficiency [75]. The rate of water vapor transport through the MD membrane is directly related to the transmembrane temperature, since the water vapor partial pressure difference (the driving force) is caused by the temperature difference at both sides of the membrane [76]. In addition, this driving force is also affected by both the temperature and feed concentration polarization effects, which are important phenomena that negatively affect the performance of MD systems. The induced concentration and temperature boundary layers at both membrane surfaces reduce the water vapor transfer by decreasing the temperature difference between the two sides of the membrane and increasing the energy consumption [77]. The fact that MD scientists frequently employ the word "performance" is proof that their main goal is to improve this parameter, as they refer to both the permeate flux and rejection factor. The improvement in MD performance over the years is also an indicator of the considerable efforts made to render this technology one of the leading membrane technologies for water production in the near future.

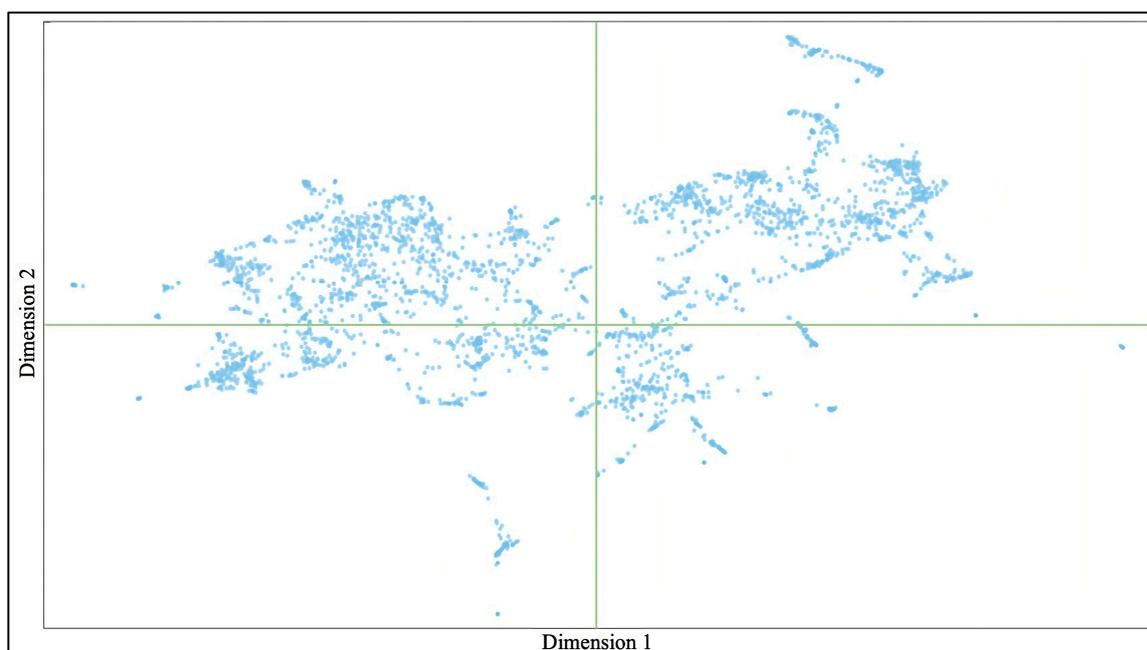


Figure 4. Representation of the collection in 2D space (*min_cluster_size* = 3684).

As can be seen in Figure 4, the documents occupied a large area in width in the two-dimensional plane, and the presence of different clusters in the form of islets was evident. Even this illustration can allow researchers to manually interpret that there may be quite many topics in the domain.

3.3. Global MD Subjects

In a second analysis, the *min_cluster_size* parameter was set to 1000, and the global concepts and their descriptive words were depicted. The resulting number of topics was two. The top 10 words for the topics and the distribution of the global themes in a 2D plane can be seen in Table 4 and Figure S1, respectively. Note that the typical behavior of the HDBSCAN approach is that it creates outliers (data points that do not fit into any topic). The HDBSCAN model creates outliers in clustering because forcing outliers into a cluster reduces the intercluster homogeneity and consistency (the BERTopic architecture aggregated the outliers together as Cluster -1). The outlier's cluster is also specified in Table 4 and highlighted in light grey in Figure S1.

Table 4. Global topics in MD domain (*min_cluster_size* = 1000).

Topic No.	Number of Papers	Topic Name
T-1 (Outliers)	410	membrane—water—distillation—process— concentration—membranes—flux— temperature—feed—using
T1	2121	membrane—water—distillation—feed—md— flux—temperature—process—heat—energy
T2	1153	membrane—membranes—surface—water— distillation—flux—PVDF—contact— MD—hydrophobic

In Table 4, one can easily understand the global subjects (i.e., the main research topics) of MD. The MD studies in Cluster 1 (T1) included efforts to reach global goals and solve the main problems of MD. The number of articles in this cluster (2121) represents the main objectives of MD researchers to increase the permeate flux, reduce the energy requirements, and prevent the temperature polarization effect. These T1 topics were already discussed above in relation to Table 3. Although the topics in the T2 set (1153 articles) were handled relatively less than T1, still they were the basic subject of MD studies. Membranes are an essential part of MD systems; thus, membrane engineering is a hot topic. MD membranes can be classified according to the membrane material (e.g., polymer or ceramic), membrane preparation technique (phase inversion or electrospinning), polymer type (polyvinylidene fluoride (PVDF), polypropylene (PP), poly(vinylidene fluoride-hexafluoropropylene) (PVDF-HFP), or polytetrafluoroethylene (PTFE)), and membrane type (flat sheet, hollow fiber, or nanofiber). Various researchers have carried out progressive studies on the preparation and modification of membranes specifically for MD [78–82]. The membrane surface modification topic can be revealed by the word surface in the T2 cluster. This is an effective method used to customize the surface of membranes for a specific application, to increase the MD performance, to minimize the wetting of membrane pores, and to reduce fouling problems, among other uses [83,84]. Thanks to the studies performed on membrane surface modification, MD applications have been extended to many wastewater types [85–88]. The term PVDF was included in the T2 cluster since this polymer, which is formed by $-(\text{CH}_2\text{CF}_2)_n-$ repeating units, exhibits excellent thermal stability, good processability, a high degree of mechanical strength, and robust chemical resistance, among other properties. In addition, PVDF can be dissolved in a variety of solvents, including N,N-dimethyl acetamide (DMAc), dimethyl formamide (DMF), and N-methyl-2-pyrrolidone (NMP) [89]. The word “contact” that appeared in T2 could refer to the MD configuration of direct contact membrane distillation and the contact angle of the membrane surface. The contact angle (θ) is a macroscopic expression of the complex interaction between a liquid and a solid surface that can provide information about the hydrophobic character and wetting of the membrane surface along with its chemistry and topography [90–92]. In this context, the water contact angle of the prepared membranes also sheds light on its rejection factor (i.e., performance). Membrane pore wetting in MD results in a decrease in the produced water quality, affecting the overall long-term stability of the membrane and its lifetime. The term “contact” may also refer to DCMD, which is the most used MD configura-

tion [14,72,93,94]. Although MD exhibits more selectivity than other membrane separation processes, the wetting phenomena is one of the drawbacks hindering the industrial potential implementation of MD technology [95,96]. Therefore, the efforts made in preparing hydrophobic or super-hydrophobic membranes are quite high, and this is the reason why the term “hydrophobic” appeared in T2. In general, as can be understood from all words appearing in the T2 clusters (such as surface, PVDF, hydrophobic, and contact), it can be confirmed that the second main hot topic of MD was membrane engineering.

3.4. Local MD Subjects

To find local clusters in the domain, the min_cluster_size value was set to 10, which meant creating a cluster if 10 or more articles contained the same topic. With this value, we aimed to maintain the stability of clusters containing very few articles at a certain level and to provide convenience in terms of interpretability. The created cluster number was 63. Note that 1173 documents were marked as outliers and did not belong to any cluster. Before proceeding to the presentation and interpretation of the generated themes, the BERTopic model provided the opportunity to fine-tune the results by revealing a similarity of concepts. The similarity matrix can be seen in Figure 5.

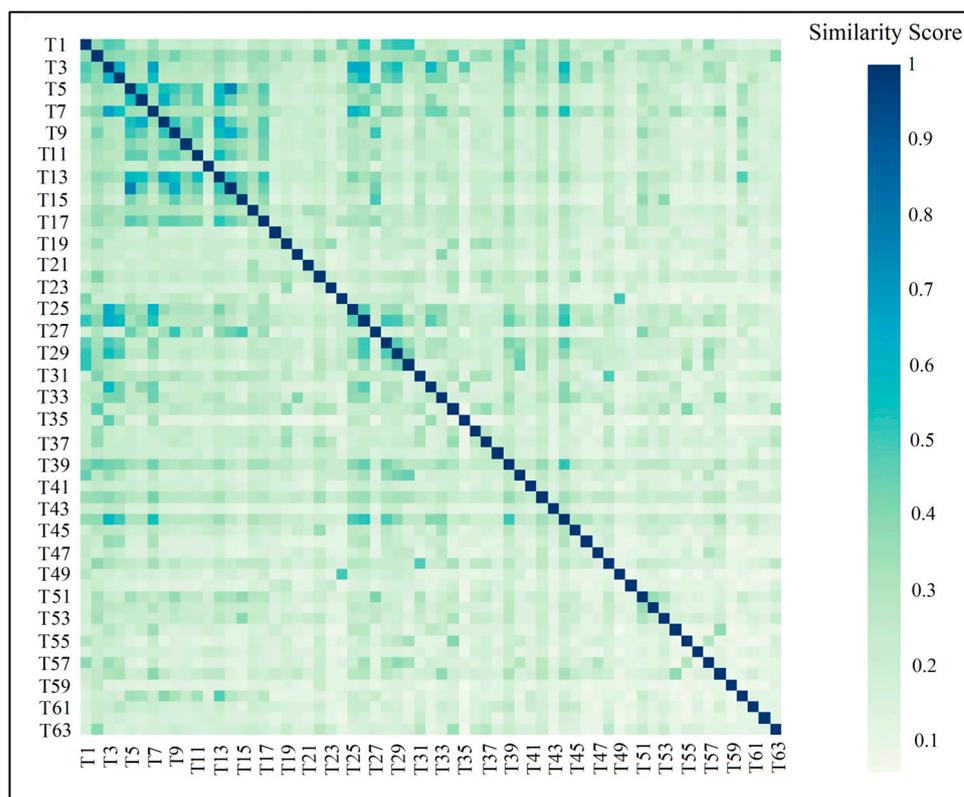


Figure 5. Similarity matrix of the resulting 63 topics.

As can be seen in Figure 5, there was a high similarity between Topic 5 (electrospun—nanofiber—nanofiber—membranes) and Topic 14 (superhydrophobic—nanofibrous—electrospun—electrospinning), with a value of ~0.76. Therefore, the results were adjusted by combining the 5th and 14th topics, and the number of clusters was reduced to 62. The resulting topics with their descriptive terms and the distribution of the topics in a 2D plane can be seen in Table 5 and Figure 6, respectively.

Table 5. Local topics in the MD domain (*min_cluster_size* = 10).

Topic No	Number of Papers	Topic Name
T1	224	solar—energy—water—desalination—production—collector—thermal—unit—plant—collectors
T2	175	scaling—crystallization—brine—crystals—scale—MD—membrane—gypsum—recovery—RO
T3	159	electrospun—nanofibrous—nanofiber—electrospinning—membranes—ENMs—
T4	121	superhydrophobic—layer—membrane—fabricated heat—transfer—module—model—mass—temperature—DCMD—flow—feed—thermal
T5	119	gap—AGMD—air—temperature—feed—flow—flux—coolant—module—permeate
T6	90	hollow—fiber—PVDF—spinning—fibers—membranes—dope—polyvinylidene—outer—inner
T7	87	gas—model—mass—vapor—flow—transport—transfer—diffusion—membrane—flux
T8	80	PVDF—membranes—phase—pore—polymer—prepared—casting—structure—porosity—properties
T9	77	superhydrophobic—surface—angle—membrane—PVDF—membranes—contact—super-hydrophobicity—modified—sliding
T10	70	carbon—CNTs—CNT—nanotube—nanotubes—CNIM—immobilized—membranes—membrane—MWCNTs
T11	67	ceramic—grafting—hydrophobic—membranes—alumina—modified—sintering—angle—contact—membrane
T12	60	juice—aroma—concentration—osmotic—fruit—compounds—apple—OMD—juices—OD
T13	59	membranes—plasma—composite—hydrophobic—membrane—hydrophilic—surface—pore—porous—prepared
T14	52	janus—oil—fouling—underwater—surface—hydrophilic—wetting—membrane—composite—hydrophobic
T15	48	dye—textile—dyeing—dyes—wastewater—disperse—reactive—permeate—color—process
T16	47	graphene—oxide—membranes—membrane—RGO—water—PVDF—surface—rejection—composite
T17	45	ammonia—pH—biogas—removal—slurry—nitrogen—NH ₃ ⁺ —CO ₂ —recovery—ammonium
T18	45	bioreactor—anaerobic—MDBR—draw—sludge—wastewater—removal—organic—OMBR/MD—OMBR
T19	41	fermentation—ethanol—broth—butanol—glucose—sugar—separation—yeast—broths—production
T20	35	photocatalytic—TiO ₂ —photocatalysis—photocatalyst—dye—degradation—PMR—catalyst—photodegradation—reactor
T21	34	acid—metals—AMD—extraction—processes—recovery—mining—treatment—pickling—pH
T22	29	FO—draw—DS—solute—FO/MD—solutes—forward—reverse—solution—hybrid
T23	29	photothermal—solar—PMD—conversion—solar-driven—efficient—light—energy—desalination—SMD
T24	28	separators—experimental—transfer—temperature—membrane—thermal—mass—PTFE—results—polarization
T25	28	fiber—water—hollow—desalination—feed—module—model—flow—rate—temperature

Table 5. Cont.

Topic No	Number of Papers	Topic Name
T26	28	omniphobic—surface—wetting—SiNPs—SDS—reentrant—tension—membrane—membranes—nanoparticles
T27	27	VMD—vacuum—feed—energy—heat—consumption—exergy—MVR—temperature—pump
T28	26	heat—cost—energy—dehumidification—pump—efficiency—cooling—MD—GOR—DCMD
T29	26	desalination—energy—technologies—RO—environmental—entropy—hybrid—generation—cost—heat
T30	25	fouling—HA—humic—vapor—pressure—layer—decline—BSA—silica—organic—flux
T31	25	fiber—module—modules—hollow—transfer—mass—CFD—baffles—flow—promoters
T32	25	ethanol—selectivity—Stefan—Maxwell—ethanol—water—feed—concentration—temperature—model—mixture—solutions
T33	24	shale—electrocoagulation—pretreatment—gas—produced—wastewater—CSG—treatment—fracking—fracturing
T34	23	spacer—spacer-filled—channels—spacers—filament—transfer—channel—CFD—heat—Reynolds
T35	22	biofilm—bacteria—biofouling—microbial—community—micropollutants—biofilms—compounds—MD—fouling
T36	22	leachate—landfill—treatment—concentrate—MD—NF—organic—wastewater—H ₂ O ₂ —AQP
T37	21	arsenic—removal—As(III)—As(V)—rejection—ppb—groundwater—contaminated—pH—Hg ⁺
T38	20	field—permeate—VMD—water—vacuum—flux—electromagnetic—feed—magnetic—salt
T39	20	OHE—power—PRMD—PRO—electricity—low-grade—exergy—heat—energy—efficiency
T40	20	radioactive—decontamination—wastes—nuclear—low-level—liquid—TeMs—waste—PET—LLRW
T41	19	process—distillation—SGMD—MD—membrane—processes—technology—sodium—review—separation
T42	19	regeneration—desiccant—regenerator—LiCl—liquid—solution—LDAC—concentration—polarisation—temperature
T43	18	feed—vacuum—VMD—temperature—flow—operating—rate—desalination—pressure—velocity
T44	17	acid—hydrochloric—concentration—HCl—sulfuric—HCl—solutions—rare—feed—earth
T45	16	ANN—neural—model—artificial—data—network—learning—accuracy—index—error
T46	16	column—separation—hybrid—membrane—distillation—processes—shortcut—design—area—propylene—optimisation
T47	15	fouling—MF—foulants—colloidal—cleaning—MD—silica—model—vibration—cake
T48	15	photothermal—NPs—plasmonic—NESMD—Ag ⁺ —NiSe—light—CoSe—conversion—solar
T49	15	urine—human—urea—diversion—nutrients—FO—nitrogen—recovery—sanitation—nutrient
T50	15	surfactant—wetting—SDS—surfactants—wetted—membrane—PAM—surface—omniphobic—Ca ²⁺

Table 5. *Cont.*

Topic No	Number of Papers	Topic Name
T51	15	wetting—detection—pore—intrusion—wetted—liquid—pressure—sucrose—distillate—Tf
T52	15	oil—oily—bilge—hexane—water—emulsion—SDS—wastewaters—nylon—produced
T53	13	chloroform—MAS—mass—VOCs—air—stripping—transfer—removal—VOC—volatile—regime
T54	12	shale—gas—fracturing—cost—management—treatment—produced—wastewater—energy—model
T55	12	OMW—olive—polyphenols—phenolic—TF200—DCMD—activity—TOW—TF1000—antioxidant
T56	12	TMD—cost—design—heat—district—optimization—MD—HEN—optimal—network
T57	11	benzene—volatile—aqueous—separation—vacuum—organic—VMD—compounds—HOVs—VOC
T58	11	lithium—Li ⁺ —extraction—brine—HMO—brines—NF—Na ⁺ —Mg ²⁺ —recovery
T59	11	PES—SMMs—blended—spectroscopy—nSMM—PET—TeMs—membranes—synthesized—contact
T60	11	boron—boric—removal—permeate—020—acid—VA—AGMD—concentration—AGMD—feed
T61	10	whey—milk—skim—lactose—dairy—IW—fouling—components—beverage—concentration
T62	10	nickel—FGDW—FGD—retentate—fouling—wastewater—desulfurization—PRO—Mg—Si—electroplating

As indicated in Table 5, there were 62 specific topics in the MD literature. Although it was not possible to interpret each topic individually in the present study, local studies of MD were most notable as solar applications (224 articles). Since MD is a thermally driven technology, interest in adopting solar-powered MD systems for desalination is expanding globally. Different types of solar systems have been successfully combined with MD, including heating with flat-plate solar collectors, heating with evacuated-tube solar collectors, heating with solar concentrators, powering with a solar pond, and photothermal collectors, among others [1,97,98]. Scaling, which is a phenomenon in which crystallization and/or precipitation of soluble salts occurs on the membrane surface [99], was included within the T2 topic of MD. Some ions in feed solutions, such as calcium and magnesium, may undergo chemical reactions to create carbonates or hydroxides, which then induce membrane scaling [100]. During a long-term MD operation, these scalants may obstruct membrane pores and eventually induce wetting, reducing the permeate flux and rejection factor as a consequence. Surface and bulk crystallization are the two processes through which mineral scalants deposit and develop on membrane surfaces [101]. Efforts have been made to overcome this important problem (175 articles). The third topic in Table 5 is nanofiber membrane fabrication via the electrospinning technique (159 articles). Electrospun nanofibrous membranes (ENMs) exhibit various advantages compared to phase-inversion membranes, such as their very high void volume fraction, high surface-to-volume ratio and hydrophobic character, and energy efficiency, among others [102,103]. Heat and mass transfer in MD are two important mechanisms affecting the produced vapor flux and thermal efficiency. Both occur simultaneously in MD systems [74,104–108]. Since DCMD is the commonly used configuration, it is expected that heat and mass transfer mechanisms are mostly investigated for this MD variant. In Table 5, it can be seen that 119 articles were grouped in the T5 topic that included AGMD, which was the second most used MD configuration, since condensation is carried out inside membrane modules over a condensing cold surface. However, due to the localized air resistance between the membrane and the condensing surface, the resulting AGMD permeate vapor flux is often minimal, although the proposed module designs with heat recovery allow a high energy efficiency [109]. The low heat transfer via conduction through the membrane fol-

lowing Fourier's law results in a low conductive heat loss through the membrane and a high thermal efficiency [110]. Polymeric hollow-fiber membranes have been used widely in most MD separation applications [111] because of their higher mechanical stability and packing density [112]. This is why this type of membrane was included in the sixth topic, with 90 articles. The two major methods considered for hollow-fiber membrane preparation were non-solvent-induced phase separation (NIPS) and thermally induced phase separation (TIPS). Because hollow-fiber membranes exhibit some unique advantages, including self-supporting (i.e., they do not require any support to withstand operation conditions) and their variety of possible arrangements in modules to achieve a high packing density and optimal fluid dynamics, reducing both the temperature and the concentration polarization effect, they attract much interest in the MD research field [113]. In general, Table 5 exhibits very useful information related to hot MD topics. This information includes polymer types (PVDF, PP, PS, and PTFE), polymer additives (carbon nanotubes (CNTs) and surface-modifying macromolecules (SMMs)), membrane configurations (VMD and SGMD), fields of application (urine, leachate, and arsenic removal), module geometry (spacer and channel), theoretical modeling (Stefan–Maxwell and CFD), artificial intelligence applications (Neural Networks and modeling), and hybridization with other separation systems (MBR and FO) investigated in MD. With the information obtained in the BERTopic modeling, MD researchers may be aware of the current status of the subject on which to work. In addition, the local topics identified through the BERTopic approach listed at the end of Table 3 can be developed in the future. Unexplored subjects should be tackled. Identifying the MD research gaps will further boost MD knowledge and advance MD technology, keeping researchers away from centralized topics. Among the research topics registered and with few published articles are heavy metals, toxic gases, and acids in wastewaters, which are major problems in industries. These topics offer great opportunities to promote and prove the versatility of MD for its industrial implementation. Topic modeling revealed the methods applied in research development (membrane type, configuration, modeling approach, etc.). For instance, DCMD is predominantly used in olive oil, polyphenols, olive mill wastewater processes (T55), while other MD configurations were not identified. In the T58 topic, the nanofiltration (NF) separation process was also involved in Li^+ recovery from brines. In this case, is it possible that any other combination with other more effective membrane separation processes such as reverse osmosis (RO), forward osmosis (FO), electrodialysis (ED), etc., will result in a better treatment efficiency or energy consumption?

Another point to be mentioned regarding Figure 6 and Table 5 is that 2511 papers were assigned to topics, but 1173 papers were highlighted as outliers and did not belong to any topic. The BERTopic architecture contains methods for reducing the number of outliers. The first approach was to adjust the `min_samples` parameter in HDBSCAN. This value was automatically set to the same value as the `min_cluster_size`. In addition, the `reduce_outliers` function in the BERTopic algorithm attempted to reduce the outliers by forcing them into a cluster. If an analyst would like to allocate every data instance to a cluster (generating no outliers), k-means can be used instead of HDBSCAN in the third step. In this study, we optimized the number of clusters. However, extra forcing of the outliers into a cluster was a bad option, since it changed the underlying structure of the data and decreased the clustering's quality. As a result, assigning them to a cluster could alter the cluster's center, shape, and size. In addition, decreasing the similarity within the cluster while increasing the similarity across clusters could lead to misleading or erroneous topics. Another approach may be to reduce the `min_cluster_size` to a value lower than 10. However, the issue in this last case is that tiny subjects were created and were so irrelevant that they could be ignored. Furthermore, a higher number of clusters could result in more complicated and less informative graphs. As a consequence, the `min_cluster_size` was kept at its optimal value. The outliers were defined with the following words: membrane—water—flux—distillation—membranes—md—feed—process—temperature—heat. These terms were similar to those in Table 3 that defined the collection but with a minor change. Instead of “performance”, the term “heat” was considered. This indicated that the

the topic and the effect of all terms in the cluster was equal, it was understood that keeping the number of descriptive terms high was more useful.

Topics over time is a statistical procedure applied to identify how a given subject in a set of documents evolves with time. It is a useful approach that helps to understand the evolution of ideas, trends, and interests. Figure S3 shows the evolution (topics over time) of the 62 local topics detected in the MD domain. The most notable situation shown in Figure S3 is that some subjects were studied with greater momentum in recent years, while others were not. By employing a simple linear regression, it was found that the top five subjects with the highest slope were T1, T2, T3, T4, and T5, with ~ 0.339 , ~ 0.254 , ~ 0.260 , ~ 0.171 , and ~ 0.170 values, respectively. Although these topics were the most researched, it was evident that interest in this research is also increasing every year. The topics with the lowest slope were T44, T53, T57, T59, and T61, with values of ~ 0.010 , ~ 0.008 , ~ 0.009 , ~ 0.015 , and ~ 0.014 , respectively. This result showed that the popularity of the last five topics has remained stable and have been studied by limited research groups over the years. T1, T6, and T23 were the earliest topics investigated in 1967, when MD was first introduced. Considering the last 5 years may be a useful way to explore the distribution and density of the topics. T1, T2, T3, T4, and T9 were the most emphasized subjects, with 115, 89, 103, 57, and 58 reported studies, respectively. The topics T44 and T53 have been ignored, since only one study corresponded to each one during the last 5 years. The number of topics with more than 25 papers published during the last 5 years was 16, while the number of topics with less than 10 papers published during the last 5 years was 23. Figure S3 also indicates that there were topics that exhibited similar patterns over time. In terms of similarity, the pairs T59–T60 and T48–T58 could be considered the closest to each other. This indicated that MD researchers sometimes have increasing or decreasing interests in different topics during the same period of time.

4. Conclusions

Membrane distillation (MD) is a promising separation technology offering appropriate solutions to the worldwide water issue by treating saline waters. This membrane process has received substantial investigation, progressing from laboratory tests and novel membrane fabrication to systems development, theoretical modeling, and optimization, making it a highly promising separation method. The present study used a state-of-the-art artificial intelligence approach—topic modeling—to uncover the main and refined topics in the MD literature. The topic modeling method has the power to provide detailed insights into a corpus by finding the themes in a very short time with substantially less effort. In this study, a dataset that included 3684 articles downloaded from the Scopus database on 23 January 2023 was analyzed using the recently developed BERTopic architecture. Depending on the results, when there was only a single cluster, the membrane research was mainly defined with the following words: “membrane”, “water”, “distillation”, “flux”, “membranes”, “MD”, “feed”, “process”, “temperature”, and “performance”. When the *min_cluster_size* parameter was set to 1000, the dataset could be divided in two clusters. Thus, two global topics together with their descriptive words could also be revealed. In one cluster, the most globally researched MD topics (2121 documents) regarded feed wastewater content, permeate flux enhancement, temperature polarization, heat, and energy requirements. The other cluster (1153 studies) was mostly focused on surface modification, PVDF polymer applications, hydrophobic membrane production, and the direct contact membrane distillation (DCMD) configuration.

The local topics were also revealed in the present study. When the *min_cluster_size* parameter was set to 10 (i.e., to create a cluster if 10 or more articles contained the same topic), 62 specialized topics were found to be investigated by researchers. Among these local topics, the most emphasized one was MD solar applications (224 research articles). The second most researched local topic regarded the membrane scaling problem and membrane crystallization applications, with 175 articles. Among the local topics, boron and boric acid removal via the air gap membrane distillation (AGMD) configuration

(11 articles), lactose removal from dairy effluents (10 articles), and elimination of some heavy metals and sulfur from electroplating wastewater (10 articles) were the least covered themes. Through the method followed in this research study, attempts were made to guide scientists who are performing research studies on MD by alerting them to the topics that have been overemphasized and/or ignored. The results of the present research may help MD researchers to have an idea regarding the topic on which they are working today and to choose the topic of their future MD scientific studies. It should be noted that since the number of scientific publications on MD is increasing exponentially every year, it is important to update the MD topic modeling analyses in the future.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/separations10090482/s1>, Figure S1: Global topics of MD domain (*min_cluster_size* = 1000); Figure S2: Topic term scores; Figure S3: Topics over time.

Author Contributions: All authors contributed equally in terms of investigation, methodology, conceptualization, and validation; E.A., software; E.A., data curation; E.A., writing—original draft; M.K., writing—review and editing; E.A., visualization; M.K., supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data and information are available upon request to the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shalaby, S.M.; Kabeel, A.E.; Abosheisha, H.F.; Elfakharany, M.K.; El-Bialy, E.; Shama, A.; Vidic, R.D. Membrane distillation driven by solar energy: A review. *J. Clean. Prod.* **2022**, *366*, 132949. [\[CrossRef\]](#)
2. Jiang, G.; Yu, W.; Lei, H. Novel solar membrane distillation system based on Ti3C2TX MXene nanofluids with high photothermal conversion efficiency. *Desalination* **2022**, *539*, 115930. [\[CrossRef\]](#)
3. Boretti, A.; Rosa, L. Reassessing the projections of the World Water Development Report. *NPJ Clean Water* **2019**, *2*, 15. [\[CrossRef\]](#)
4. Khayet, M.; Matsuura, T. *Membrane Distillation-Principles and Applications*; Elsevier: Amsterdam, The Netherlands, 2011.
5. Ansari, A.; Kavousi, S.; Helfer, F.; Millar, G.; Thiel, D.V. An Improved Modelling Approach for the Comprehensive Study of Direct Contact Membrane Distillation. *Membranes* **2021**, *11*, 308. [\[CrossRef\]](#)
6. Chen, H.; Mao, Y.; Mo, B.; Pan, Y.; Xu, R.; Ji, W.; Chen, G.; Liu, G.; Jin, W. Plasma-assisted facile fabrication of omniphobic graphene oxide membrane with anti-wetting property for membrane distillation. *J. Membr. Sci.* **2023**, *668*, 121207. [\[CrossRef\]](#)
7. Han, L.; Mao, J.; Xie, A.-Q.; Liang, Y.; Zhu, L.; Chen, S. Synergistic enhanced solar-driven water purification and CO₂ reduction via photothermal catalytic membrane distillation. *Sep. Purif. Technol.* **2023**, *309*, 123003. [\[CrossRef\]](#)
8. Ju, J.; Huang, Y.; Liu, M.; Xie, N.; Shi, J.; Fan, Y.; Zhao, Y.; Kang, W. Construction of electrospinning Janus nanofiber membranes for efficient solar-driven membrane distillation. *Sep. Purif. Technol.* **2023**, *305*, 122348. [\[CrossRef\]](#)
9. Shafieian, A.; Khiadani, M.; Zargar, M. Performance analysis of tubular membrane distillation modules: An experimental and CFD analysis. *Chem. Eng. Res. Des.* **2022**, *183*, 478–493. [\[CrossRef\]](#)
10. Liu, Z.; Lu, X.; Zhang, S.; Ma, R.; Gu, J.; Ren, K.; Liu, C. Study on the synergistic heat transfer of double boundary layers in the jacketed vacuum membrane distillation process. *Desalination* **2023**, *549*, 116356. [\[CrossRef\]](#)
11. Mutlu-Salmanli, O.; Eryildiz, B.; Vatanpour, V.; Deliballi, Z.; Kiskan, B.; Koyuncu, I. Fabrication of novel hydrophobic electrospun nanofiber membrane using polybenzoxazine for membrane distillation application. *Desalination* **2023**, *546*, 116203. [\[CrossRef\]](#)
12. Lin, J.; Du, J.; Xie, S.; Yu, F.; Fang, S.; Yan, Z.; Lin, X.; Zou, D.; Xie, M.; Ye, W. Durable superhydrophobic polyvinylidene fluoride membranes via facile spray-coating for effective membrane distillation. *Desalination* **2022**, *538*, 115925. [\[CrossRef\]](#)
13. Zhang, Y.; Ji, Z.; Yan, H.; Wu, B.; Guo, Y.; Wang, H.; Li, C. Water recovery from cleaning wastewater of traditional Chinese medicine processing via vacuum membrane distillation: Parameters optimization and membrane fouling investigation. *Chem. Eng. Res. Des.* **2022**, *188*, 555–563. [\[CrossRef\]](#)
14. Aytac, E.; Khayet, M. A deep dive into membrane distillation literature with data analysis, bibliometric methods, and machine learning. *Desalination* **2023**, *553*, 116482. [\[CrossRef\]](#)
15. Acevedo, L.; Uche, J.; Del-Amo, A. Improving the Distillate Prediction of a Membrane Distillation Unit in a Trigeration Scheme by Using Artificial Neural Networks. *Water* **2018**, *10*, 310. [\[CrossRef\]](#)
16. Dudchenko, A.V.; Mauter, M.S. Neural networks for estimating physical parameters in membrane distillation. *J. Membr. Sci.* **2020**, *610*, 118285. [\[CrossRef\]](#)
17. Chamani, H.; Yazgan-Birgi, P.; Matsuura, T.; Rana, D.; Hassan Ali, M.I.; Arafat, H.A.; Lan, C.Q. CFD-based genetic programming model for liquid entry pressure estimation of hydrophobic membranes. *Desalination* **2020**, *476*, 114231. [\[CrossRef\]](#)

18. Huang, J.; Tang, T.; He, Y. Numerical Simulation Study on the Mass and Heat Transfer in the Self-Heating Membrane Distillation Process. *Ind. Eng. Chem. Res.* **2021**, *60*, 12663–12674. [[CrossRef](#)]
19. Ali, K.; Arafat, H.A.; Hassan Ali, M.I. Detailed numerical analysis of air gap membrane distillation performance using different membrane materials and porosity. *Desalination* **2023**, *551*, 116436. [[CrossRef](#)]
20. Adel, K.; Elhakeem, A.; Marzouk, M. Decentralizing construction AI applications using blockchain technology. *Expert Syst. Appl.* **2022**, *194*, 116548. [[CrossRef](#)]
21. Aytac, E. Modeling Future Impacts on Land Cover of Rapid Expansion of Hazelnut Orchards: A Case Study on Samsun, Turkey. *Eur. J. Sust. Dev. Res.* **2022**, *6*, em0193. [[CrossRef](#)]
22. Habuza, T.; Navaz, A.N.; Hashim, F.; Alnajjar, F.; Zaki, N.; Serhani, M.A.; Statsenko, Y. AI applications in robotics, diagnostic image analysis and precision medicine: Current limitations, future trends, guidelines on CAD systems for medicine. *Inform. Med. Unlocked* **2021**, *24*, 100596. [[CrossRef](#)]
23. Aytac, E. Exploring Electrocoagulation Through Data Analysis And Text Mining Perspectives. *Environ. Eng. Man. J.* **2022**, *21*, 671–685.
24. Van Giffen, B.; Herhausen, D.; Fahse, T. Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods. *J. Bus. Res.* **2022**, *144*, 93–106. [[CrossRef](#)]
25. Aytac, E. Forecasting Turkey's Hazelnut Export Quantities with Facebook's Prophet Algorithm and Box-Cox Transformation. *ADCAIJ Adv. Distrib. Comput. Artif. Intell. J.* **2021**, *10*, 33–47. [[CrossRef](#)]
26. Houssein, E.H.; Abohashima, Z.; Elhoseny, M.; Mohamed, W.M. Machine learning in the quantum realm: The state-of-the-art, challenges, and future vision. *Expert Syst. Appl.* **2022**, *194*, 116512. [[CrossRef](#)]
27. Aytac, E. Havzaların Benzerliklerini Tanımlamada Alternatif Bir Yaklaşım: Hiyerarşik Kümeleme Yöntemi Uygulaması. *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilim. Derg.* **2021**, *21*, 958–970. [[CrossRef](#)]
28. Bao, J.; Chen, Y.; Yin, J.; Chen, X.; Zhu, D. Exploring topics and trends in Chinese ATC incident reports using a domain-knowledge driven topic model. *J. Air Transp. Manag.* **2023**, *108*, 102374. [[CrossRef](#)]
29. Abdelrazek, A.; Eid, Y.; Gawish, E.; Medhat, W.; Hassan, A. Topic modeling algorithms and applications: A survey. *Inf. Syst.* **2023**, *112*, 102131. [[CrossRef](#)]
30. Wang, F.; Zhou, R.; Feng, Y.; Lu, X. Bayesian sparse joint dynamic topic model with flexible lead-lag order. *Inf. Sci.* **2022**, *616*, 392–410. [[CrossRef](#)]
31. Gencoglu, B.; Helms-Lorenz, M.; Maulana, R.; Jansen, E.P.W.A.; Gencoglu, O. Machine and expert judgments of student perceptions of teaching behavior in secondary education: Added value of topic modeling with big data. *Comput. Educ.* **2023**, *193*, 104682. [[CrossRef](#)]
32. Feng, J.; Zhang, Z.; Ding, C.; Rao, Y.; Xie, H.; Wang, F.L. Context reinforced neural topic modeling over short texts. *Inf. Sci.* **2022**, *607*, 79–91. [[CrossRef](#)]
33. Zhang, K.; Lin, N.; Tian, G.; Yang, J.; Wang, D.; Jin, Z. Unsupervised-learning based self-organizing neural network using multi-component seismic data: Application to Xujiahe tight-sand gas reservoir in China. *J. Pet. Sci. Eng.* **2022**, *209*, 109964. [[CrossRef](#)]
34. Su, H.; Yang, X.; Xiang, L.; Hu, A.; Xu, Y. A novel method based on deep transfer unsupervised learning network for bearing fault diagnosis under variable working condition of unequal quantity. *Knowl.-Based Syst.* **2022**, *242*, 108381. [[CrossRef](#)]
35. Dehghani, M.; Ebrahimi, F. ParsBERT topic modeling of Persian scientific articles about COVID-19. *Inform. Med. Unlocked* **2023**, *36*, 101144. [[CrossRef](#)] [[PubMed](#)]
36. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794. [[CrossRef](#)]
37. Jeon, E.; Yoon, N.; Sohn, S.Y. Exploring new digital therapeutics technologies for psychiatric disorders using BERTopic and PatentSBERTa. *Technol. Forecast. Soc. Change* **2023**, *186*, 122130. [[CrossRef](#)]
38. Saidi, F.; Trabelsi, Z.; Thangaraj, E. A novel framework for semantic classification of cyber terrorist communities on Twitter. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105271. [[CrossRef](#)]
39. Grootendorst, M. BERTopic Web Page. Available online: <https://maartengr.github.io/BERTopic/index.html> (accessed on 27 January 2023).
40. López-Santillán, R.; Montes-Y-Gómez, M.; González-Gurrola, L.C.; Ramírez-Alonso, G.; Prieto-Ordaz, O. Richer Document Embeddings for Author Profiling tasks based on a heuristic search. *Inf. Process. Manag.* **2020**, *57*, 102227. [[CrossRef](#)]
41. Rahimi, Z.; Homayounpour, M.M. Tens-embedding: A Tensor-based document embedding method. *Expert Syst. Appl.* **2020**, *162*, 113770. [[CrossRef](#)]
42. Wei, C.; Luo, S.; Guo, J.; Wu, Z.; Pan, L. Discriminative locally document embedding: Learning a smooth affine map by approximation of the probabilistic generative structure of subspace. *Knowl.-Based Syst.* **2017**, *121*, 41–57. [[CrossRef](#)]
43. Singh, K.N.; Devi, S.D.; Devi, H.M.; Mahanta, A.K. A novel approach for dimension reduction using word embedding: An enhanced text classification approach. *Int. J. Inf. Manag. Data Insights* **2022**, *2*, 100061. [[CrossRef](#)]
44. Zhou, R.; Gao, W.; Ding, D.; Liu, W. Supervised dimensionality reduction technology of generalized discriminant component analysis and its kernelization forms. *Pattern Recognit.* **2022**, *124*, 108450. [[CrossRef](#)]
45. Bibal, A.; Clarinval, A.; Dumas, B.; Fréney, B. IXVC: An interactive pipeline for explaining visual clusters in dimensionality reduction visualizations with decision trees. *Array* **2021**, *11*, 100080. [[CrossRef](#)]

46. Wang, S.; Bai, L.; Chen, X.; Wang, Z.; Shao, Y.-H. Divergent Projection Analysis for Unsupervised Dimensionality Reduction. *Procedia Comput. Sci.* **2022**, *199*, 384–391. [CrossRef]
47. Ao, Z.; Horváth, G.; Sheng, C.; Song, Y.; Sun, Y. Skill requirements in job advertisements: A comparison of skill-categorization methods based on wage regressions. *Inf. Process. Manag.* **2023**, *60*, 103185. [CrossRef]
48. Ezugwu, A.E.; Ikotun, A.M.; Oyelade, O.O.; Abualigah, L.; Agushaka, J.O.; Eke, C.I.; Akinyelu, A.A. A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Eng. Appl. Artif. Intell.* **2022**, *110*, 104743. [CrossRef]
49. Aytac, E. Unsupervised learning approach in defining the similarity of catchments: Hydrological response unit based k-means clustering, a demonstration on Western Black Sea Region of Turkey. *Int. Soil Water Conserv. Res.* **2020**, *8*, 321–331. [CrossRef]
50. Ghasemi, Z.; Khorshidi, H.A.; Aickelin, U. Multi-objective Semi-supervised clustering for finding predictive clusters. *Expert Syst. Appl.* **2022**, *195*, 116551. [CrossRef]
51. Zhou, Z.; Si, G.; Sun, H.; Qu, K.; Hou, W. A robust clustering algorithm based on the identification of core points and KNN kernel density estimation. *Expert Syst. Appl.* **2022**, *195*, 116573. [CrossRef]
52. Zhu, M.-X.; Lv, X.-J.; Chen, W.-J.; Li, C.-N.; Shao, Y.-H. Local density peaks clustering with small size distance matrix. *Procedia Comput. Sci.* **2022**, *199*, 331–338. [CrossRef]
53. Khayet, M.; Aytac, E.; Matsuura, T. Bibliometric and sentiment analysis with machine learning on the scientific contribution of Professor Srinivasa Sourirajan. *Desalination* **2022**, *543*, 116095. [CrossRef]
54. Zheng, W.; Gao, J.; Wu, X.; Liu, F.; Xun, Y.; Liu, G.; Chen, X. The impact factors on the performance of machine learning-based vulnerability detection: A comparative study. *J. Syst. Softw.* **2020**, *168*, 110659. [CrossRef]
55. Qorib, M.; Oladunni, T.; Denis, M.; Ososanya, E.; Cotae, P. Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. *Expert Syst. Appl.* **2023**, *212*, 118715. [CrossRef] [PubMed]
56. Findley, M.E. Vaporization through Porous Membranes. *Ind. Eng. Chem. Process Des. Dev.* **1967**, *6*, 226–230. [CrossRef]
57. Shome, S.; Hassan, M.K.; Verma, S.; Panigrahi, T.R. Impact investment for sustainable development: A bibliometric analysis. *Int. Rev. Econ. Finance* **2023**, *84*, 770–800. [CrossRef]
58. Ng, J.Y.; Chiong, J.D.; Liu, M.Y.M.; Pang, K.K.Y. Characteristics of the Echinacea Spp. research literature: A bibliometric analysis. *Eur. J. Integr. Med.* **2023**, *57*, 102216. [CrossRef]
59. SBERT. Pretrained Models. Available online: https://www.sbert.net/docs/pretrained_models.html (accessed on 28 January 2023).
60. Yang, D.; Wei, V.; Jin, Z.; Yang, Z.; Chen, X. A UMAP-based clustering method for multi-scale damage analysis of laminates. *Appl. Math. Model.* **2022**, *111*, 78–93. [CrossRef]
61. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2018**, arXiv:1802.03426. [CrossRef]
62. Leonard, A.; Wheeler, S.; McCulloch, M. Power to the people: Applying citizen science and computer vision to home mapping for rural energy access. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102748. [CrossRef]
63. Campello, R.J.G.B.; Moulavi, D.; Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. In Proceedings of the Advances in Knowledge Discovery and Data Mining, Gold Coast, Australia, 14–17 April 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 160–172.
64. Ghamarian, I.; Marquis, E.A. Hierarchical density-based cluster analysis framework for atom probe tomography data. *Ultramicroscopy* **2019**, *200*, 28–38. [CrossRef]
65. Satrya, W.F.; Aprilliyani, R.; Yossy, E.H. Sentiment analysis of Indonesian police chief using multi-level ensemble model. *Procedia Comput. Sci.* **2023**, *216*, 620–629. [CrossRef]
66. Li, Q.; Jin, Z.; Wang, C.; Zeng, D.D. Mining opinion summarizations using convolutional neural networks in Chinese microblogging systems. *Knowl.-Based Syst.* **2016**, *107*, 289–300. [CrossRef]
67. Aytac, E.; Fombona-Pascual, A.; Lado, J.J.; Quismondo, E.G.; Palma, J.; Khayet, M. Faradaic deionization technology: Insights from bibliometric, data mining and machine learning approaches. *Desalination* **2023**, *563*, 116715. [CrossRef]
68. Grootendorst, M. BERTopic Hyperparameter Tuning. Available online: https://maartengr.github.io/BERTopic/getting_started/parameter%20tuning/parametertuning.html (accessed on 15 March 2023).
69. McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.* **2017**, *2*, 205. [CrossRef]
70. Guo, X.; Su, Q.-W.; Zhang, L.-Z. The measurement of permeate flux based on a noninvasive method for membrane distillation: Experiment and model validation. *Int. J. Heat Mass Transfer* **2021**, *164*, 120482. [CrossRef]
71. Mustafa, I.; Kilibay, A.; Alhseinat, E.; Almarzooqi, F. Enhanced Membrane Distillation Water Flux through Electromagnetism. *Chem. Eng. Process. Process Intensif.* **2021**, *169*, 108597. [CrossRef]
72. Parani, S.; Oluwafemi, O.S. Membrane Distillation: Recent Configurations, Membrane Surface Engineering, and Applications. *Membranes* **2021**, *11*, 934. [CrossRef]
73. Mohammad Reza Shirzad, K.; Ahmad, R. Membrane Distillation: Basics, Advances, and Applications. In *Advances in Membrane Technologies*; Amira, A., Ed.; IntechOpen: Rijeka, Croatia, 2020; Chapter 4.
74. Chen, L.; Xu, P.; Wang, H. Interplay of the Factors Affecting Water Flux and Salt Rejection in Membrane Distillation: A State-of-the-Art Critical Review. *Water* **2020**, *12*, 2841. [CrossRef]
75. Ullah, R.; Khraisheh, M.; Esteves, R.J.; McLeskey, J.T.; AlGhouthi, M.; Gad-el-Hak, M.; Vahedi Tafreshi, H. Energy efficiency of direct contact membrane distillation. *Desalination* **2018**, *433*, 56–67. [CrossRef]

76. Belessiotis, V.; Kalogirou, S.; Delyannis, E. Chapter Four—Membrane Distillation. In *Thermal Solar Desalination*; Belessiotis, V., Kalogirou, S., Delyannis, E., Eds.; Academic Press: Cambridge, MA, USA, 2016; pp. 191–251.
77. Anvari, A.; Azimi Yancheshme, A.; Kekre, K.M.; Ronen, A. State-of-the-art methods for overcoming temperature polarization in membrane distillation process: A review. *J. Membr. Sci.* **2020**, *616*, 118413. [[CrossRef](#)]
78. Mortaheb, H.R.; Baghban Salehi, M.; Rajabzadeh, M. Optimized hybrid PVDF/graphene membranes for enhancing performance of AGMD process in water desalination. *J. Ind. Eng. Chem.* **2021**, *99*, 407–421. [[CrossRef](#)]
79. Li, M.; Lu, K.J.; Wang, L.; Zhang, X.; Chung, T.-S. Janus membranes with asymmetric wettability via a layer-by-layer coating strategy for robust membrane distillation. *J. Membr. Sci.* **2020**, *603*, 118031. [[CrossRef](#)]
80. Ursino, C.; Di Nicolò, E.; Gabriele, B.; Criscuoli, A.; Figoli, A. Development of a novel perfluoropolyether (PFPE) hydrophobic/hydrophilic coated membranes for water treatment. *J. Membr. Sci.* **2019**, *581*, 58–71. [[CrossRef](#)]
81. Li, X.; Deng, L.; Yu, X.; Wang, M.; Wang, X.; García-Payo, C.; Khayet, M. A novel profiled core-shell nanofibrous membrane for wastewater treatment by direct contact membrane distillation. *J. Mater. Chem. A* **2016**, *4*, 14453–14463. [[CrossRef](#)]
82. Peng, Y.; Dong, Y.; Fan, H.; Chen, P.; Li, Z.; Jiang, Q. Preparation of polysulfone membranes via vapor-induced phase separation and simulation of direct-contact membrane distillation by measuring hydrophobic layer thickness. *Desalination* **2013**, *316*, 53–66. [[CrossRef](#)]
83. Chew, N.G.P.; Zhao, S.; Malde, C.; Wang, R. Superoleophobic surface modification for robust membrane distillation performance. *J. Membr. Sci.* **2017**, *541*, 162–173. [[CrossRef](#)]
84. Madalosso, H.B.; Machado, R.; Hotza, D.; Marangoni, C. Membrane Surface Modification by Electrospinning, Coating, and Plasma for Membrane Distillation Applications: A State-of-the-Art Review. *Adv. Eng. Mater.* **2021**, *23*, 2001456. [[CrossRef](#)]
85. Gryta, M. Surface modification of polypropylene membrane by helium plasma treatment for membrane distillation. *J. Membr. Sci.* **2021**, *628*, 119265. [[CrossRef](#)]
86. Huang, Y.-X.; Liang, D.-Q.; Luo, C.-H.; Zhang, Y.; Meng, F. Liquid-like surface modification for effective anti-scaling membrane distillation with uncompromised flux. *J. Membr. Sci.* **2021**, *637*, 119673. [[CrossRef](#)]
87. Hendren, Z.D.; Brant, J.; Wiesner, M.R. Surface modification of nanostructured ceramic membranes for direct contact membrane distillation. *J. Membr. Sci.* **2009**, *331*, 1–10. [[CrossRef](#)]
88. Zuo, G.; Wang, R. Novel membrane surface modification to enhance anti-oil fouling property for membrane distillation application. *J. Membr. Sci.* **2013**, *447*, 26–35. [[CrossRef](#)]
89. Kang, G.-d.; Cao, Y.-m. Application and modification of poly(vinylidene fluoride) (PVDF) membranes—A review. *J. Membr. Sci.* **2014**, *463*, 145–165. [[CrossRef](#)]
90. Hebbar, R.S.; Isloor, A.M.; Ismail, A.F. Chapter 12—Contact Angle Measurements. In *Membrane Characterization*; Hilal, N., Ismail, A.F., Matsuura, T., Oatley-Radcliffe, D., Eds.; Elsevier: Amsterdam, The Netherlands, 2017; pp. 219–255.
91. Akbari, R.; Antonini, C. Contact angle measurements: From existing methods to an open-source tool. *Adv. Colloid Interface Sci.* **2021**, *294*, 102470. [[CrossRef](#)]
92. Warsinger, D.M.; Servi, A.; Connors, G.B.; Mavukkandy, M.O.; Arafat, H.A.; Gleason, K.K.; Lienhard, V.J.H. Reversing membrane wetting in membrane distillation: Comparing dryout to backwashing with pressurized air. *Environ. Sci. Water Res. Technol.* **2017**, *3*, 930–939. [[CrossRef](#)]
93. Ismail, M.S.; Mohamed, A.M.; Poggio, D.; Pourkashanian, M. Direct contact membrane distillation: A sensitivity analysis and an outlook on membrane effective thermal conductivity. *J. Membr. Sci.* **2021**, *624*, 119035. [[CrossRef](#)]
94. Ashoor, B.B.; Mansour, S.; Giwa, A.; Dufour, V.; Hasan, S.W. Principles and applications of direct contact membrane distillation (DCMD): A comprehensive review. *Desalination* **2016**, *398*, 222–246. [[CrossRef](#)]
95. Wae AbdulKadir, W.A.F.; Ahmad, A.L.; Seng, O.B.; Che Lah, N.F. Biomimetic hydrophobic membrane: A review of anti-wetting properties as a potential factor in membrane development for membrane distillation (MD). *J. Ind. Eng. Chem.* **2020**, *91*, 15–36. [[CrossRef](#)]
96. Goh, P.S.; Naim, R.; Rahbari-Sisakht, M.; Ismail, A.F. Modification of membrane hydrophobicity in membrane contactors for environmental remediation. *Sep. Purif. Technol.* **2019**, *227*, 115721. [[CrossRef](#)]
97. Qtaishat, M.R.; Banat, F. Desalination by solar powered membrane distillation systems. *Desalination* **2013**, *308*, 186–197. [[CrossRef](#)]
98. Ma, Q.; Xu, Z.; Wang, R. Distributed solar desalination by membrane distillation: Current status and future perspectives. *Water Res.* **2021**, *198*, 117154. [[CrossRef](#)] [[PubMed](#)]
99. Gryta, M. Calcium sulphate scaling in membrane distillation process. *Chem. Pap.* **2009**, *63*, 146–151. [[CrossRef](#)]
100. Gryta, M. Alkaline scaling in the membrane distillation process. *Desalination* **2008**, *228*, 128–134. [[CrossRef](#)]
101. Liao, X.; Chou, S.; Gu, C.; Zhang, X.; Shi, M.; You, X.; Liao, Y.; Razaqpur, A.G. Engineering omniphobic corrugated membranes for scaling mitigation in membrane distillation. *J. Membr. Sci.* **2023**, *665*, 121130. [[CrossRef](#)]
102. Ma, H.; Hsiao, B.S. Chapter 4—Electrospun Nanofibrous Membranes for Desalination. In *Current Trends and Future Developments on (Bio-) Membranes*; Basile, A., Curcio, E., Inamuddin, Eds.; Elsevier: Amsterdam, The Netherlands, 2019; pp. 81–104.
103. Subrahmanya, T.M.; Arshad, A.B.; Lin, P.T.; Widakdo, J.; Makari, H.K.; Austria, H.F.M.; Hu, C.-C.; Lai, J.-Y.; Hung, W.-S. A review of recent progress in polymeric electrospun nanofiber membranes in addressing safe water global issues. *RSC Adv.* **2021**, *11*, 9638–9663. [[CrossRef](#)]
104. Mengual, J.I.; Khayet, M.; Godino, M.P. Heat and mass transfer in vacuum membrane distillation. *Int. J. Heat Mass Transf.* **2004**, *47*, 865–875. [[CrossRef](#)]

105. Orfi, J.; Loussif, N.; Davies, P.A. Heat and mass transfer in membrane distillation used for desalination with slip flow. *Desalination* **2016**, *381*, 135–142. [[CrossRef](#)]
106. Bandini, S.; Gostoli, C.; Sarti, G.C. Role of heat and mass transfer in membrane distillation process. *Desalination* **1991**, *81*, 91–106. [[CrossRef](#)]
107. Xu, L.; Xu, S.; Wu, X.; Wang, P.; Jin, D.; Hu, J.; Li, L.; Chen, L.; Leng, Q.; Wu, D. Heat and mass transfer evaluation of air-gap diffusion distillation by ϵ -NTU method. *Desalination* **2020**, *478*, 114281. [[CrossRef](#)]
108. Qtaishat, M.; Matsuura, T.; Kruczek, B.; Khayet, M. Heat and mass transfer analysis in direct contact membrane distillation. *Desalination* **2008**, *219*, 272–292. [[CrossRef](#)]
109. Alkhudhiri, A.; Hilal, N. 3-Membrane distillation—Principles, applications, configurations, design, and implementation. In *Emerging Technologies for Sustainable Desalination Handbook*; Gude, V.G., Ed.; Butterworth-Heinemann: Oxford, UK, 2018; pp. 55–106.
110. Khayet, M.; Cojocaru, C. Air gap membrane distillation: Desalination, modeling and optimization. *Desalination* **2012**, *287*, 138–145. [[CrossRef](#)]
111. Wan, C.F.; Yang, T.; Lipscomb, G.G.; Stookey, D.J.; Chung, T.-S. Design and fabrication of hollow fiber membrane modules. *J. Membr. Sci.* **2017**, *538*, 96–107. [[CrossRef](#)]
112. Lu, K.-J.; Wang, P.; Chung, T.-S. Chapter 23—Hollow fiber membranes for membrane distillation applications. In *Hollow Fiber Membranes*; Chung, T.-S., Feng, Y., Eds.; Elsevier: Amsterdam, The Netherlands, 2021; pp. 495–521.
113. Pagliero, M.; Khayet, M.; García-Payo, C.; García-Fernández, L. Hollow fibre polymeric membranes for desalination by membrane distillation technology: A review of different morphological structures and key strategic improvements. *Desalination* **2021**, *516*, 115235. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.