Nigerian Research Journal of Engineering and Environmental Sciences 8(1) 2023 pp. 92-101
p ISSN: 2635-3342; e ISSN: 2635-3350

Nigerian Research Journal of Engineering and Environmental Sciences
Journal homepage: www.rjees.com

Original Research Article

Towards the Development of Machine Learning Models for Deep Cycle Battery Performance Analysis: Research Gap Identification via Bibliometric Analysis and Global Evidence Mapping

***Obayuwana, A. and Okoye, J.U.**

Department of Computer Engineering, Faculty of Engineering, University of Benin, PMB 1154, Benin City, Nigeria.
*augustine.obayuwana@uniben.edu; jokoye360@gmail.com

http://doi.org/10.5281/zenodo.8094134

ARTICLE INFORMATION

Article history:
Received 21 Apr. 2023
Revised 16 May 2023
Accepted 19 May 2023
Available online 30 Jun. 2023

Keywords:
Deep cycle battery
Machine learning
Bibliometric analysis
Research gap identification
Scopus review
VOSviewer

ABSTRACT

The use of machine learning has recently increased in the field of energy storage. This is due to the level of accuracy and convenience in estimating battery states like the state of health (SOH), state of charge (SOC), and remaining useful life (RUL) which is used to estimate the performance of the battery. A lot of researchers have used machine models to estimate these battery states. These models each differ in approach, outcome, and effectiveness. This study was undertaken to identify research gaps in topics related to machine learning models for battery performance analysis. These related works were identified using bibliometric analysis coupled with a systematic literature review of the study search index through the Scopus database-indexed publications. The results from this study show 286 articles selected from 3,086 documents identified via a visualization of the network map rendered using VOSviewer. Sensitivity analysis, cycle life, and lead acid batteries were revealed as the particular research gaps. The study also presented results of mapping different methods used to identify, prioritize, and visualize research gaps. The study output would prove a useful resource for quick insight into the techniques and methodologies in this field.

1. INTRODUCTION

Machine learning (ML) is the study of computer programs that leverage algorithms and statistical models using provided data to learn through inference and patterns without being explicitly programmed (Boutaba, et al, 2018). Machine learning algorithms make use of sample data to build models that can be used to make decisions or predictions intelligently. Machine learning, being a subset of artificial intelligence as shown in Figure 1, heavily uses statistical methods and is dependent on the quality of its dataset for acceptable results (Sindhu et al, 2020).

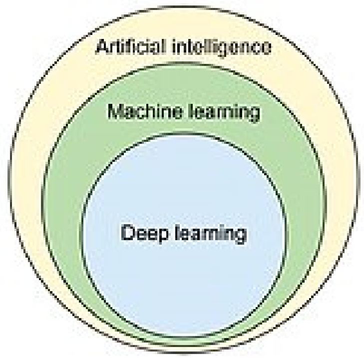

Figure 1: ML as a subfield of AI (Sindhu et al., 2020)

Machine learning models are programs that make use of an algorithm to find patterns in a dataset and consequently make decisions (Obayuwana and Monica, 2021). Since the term "Machine learning" was coined by Arthur Samuel in 1952, there have been developments and innovations in this field.

In more recent years, there have been more advanced usage of ML such as the IBM Watson which beat two champions in the Jeopardy game show in 2011, Google's Brain which teaches itself to recognize cats through Youtube Videos, Facebook's DeepFace algorithm capable of identifying individuals, and Google's AlphaGo program that beat a professional human Go player in 2016 (Sindhu, et al, 2020). A new era of neural networks has developed called Deep learning. This phrase simply refers to neural network models with many wide successive layers. ML is now responsible for most of the significant improvements in technology (Lavecchia, 2019). ML models are becoming more adaptive owing to the fact that they become more accurate the longer they operate.

The knowledge of machine learning has been increasing in recent years and it is finding applications in so many technological innovations. Integrating machine learning and human knowledge has facilitated the interaction between human beings and machine-learning systems, making machine-learning decisions understandable to humans (Chen et al 2020). Machine Learning has found its applications in areas such as natural language processing, image recognition, and classification, self-driving cars, financial fraud detection, robotics, etc (Issame, 2022).

Outside the field of data sciences and computer science, machine learning finds its application in the field of chemistry (Goh et al., 2017), engineering, (Flah et al., 2021), physics (Dunjko and Briegel, 2018), and materials science (Morgan and Jacobs, 2020). Besides the above-mentioned disciplines, machine learning technologies have great potential for addressing the improvement and management of energy storage devices and systems by significantly improving prediction accuracy and computational efficiency. Due to this, a lot of researchers have embraced the use of data science rather than traditional numerical analysis. An excellent study in this field is to estimate and predict the health and life cycle of batteries.

A battery can support only a finite, limited number of charge and discharge cycles. Thus, improper use deteriorates the battery's performance and eventually shortens its life span. Therefore, a battery management system (BMS) is of great importance in order to ensure that a battery is operated within its specified safety limits. The nonlinear complex behavior of the battery makes estimating battery SOC and SOH very challenging, yet necessary.

Battery life cycle prediction uses numerical methods such as data-driven support vector machines (SVM) or Gaussian process regression in the early days (Laayouj et al., 2016). However, these methods still do not fully describe the nonlinear battery characteristic curve, so the combination of these methods makes it more complete. Some data features are periodic, some have specific amplitudes, and some have specific conditions. This is a considerable challenge in the past, but machine learning can easily solve these problems. Long short-term memory (LSTM) and Recurrent Neural Networks (RNN) are often used in timing-related parameter analysis (Wang, et al, 2020). Deep Neural Networks (DNN) analyze more parameters with more complex variables (Khumprom and Yodo, 2019).

The availability of numerous machine learning models for estimating battery SOH and SOC makes the choice of an efficient model for researchers and manufacturers a difficult job.

2. METHODOLOGY

2.1. Research Process

The use of bibliometric analysis to identify research gaps in various fields begins with a search of the selected database to identify related documents. The search results are arranged and filtered to remove irrelevant literature. The resulting data is then inputted into appropriate software for visualization. Figure 2 shows the process used to complete this project.

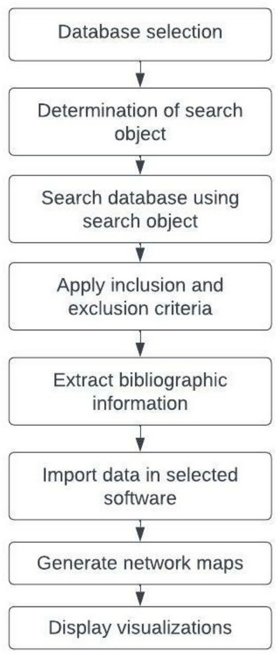

Figure 2: Workflow for the project

2.2. Identification of Documents Related to the Study

To identify gaps in a field of study, a proper analysis of published literature is carried out. Due to the large amount of literature available, there has to be a systematic selection of the literature and publications which will be included in the study. Proper identification of these publications has to be made to ensure that only the publications relating to the area of study are considered. The method used for this selection is explained in the next section.

2.3. Database Selection

For the analysis, a database is needed to enable the extraction of bibliographic data. The database contains a wide collection of literature and citations and other metrics which are important to the literary work such as publication year, publication type, keywords, and field of study. The popular databases that can be used for extraction of bibliographic work include the Scopus database and Web of Science; while others include Dimensions database, PubMed, etc. The Scopus database was used for this project. The Scopus database is an abstract and citation database that is used for research and analysis. With about 84 million records, it has one of the widest coverages of literature. It is an important tool in bibliometric analysis because it provides a wide range of literature for a related topic or keyword. Figure 3 shows an overview of the homepage of Scopus.

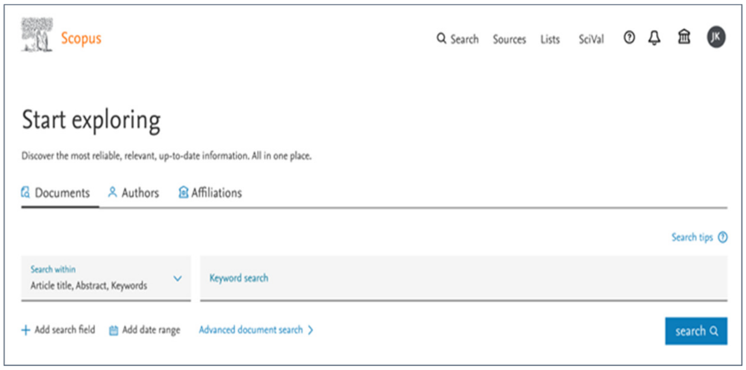

Figure 3: Overview of Scopus database home page

2.4. Determination of Search Objects and Time-span

The correct search item is important in the selection of documents. Scopus basic search was used for the work. The term "Machine Learning" yielded about 495,787 results. More streamlining was done by adding the word (AND "Battery") to indicate papers that used machine learning for battery applications. The search yielded 3,086 results. Searches can be improved by using Boolean operators (AND, OR, AND NOT). The string ("Machine Learning" AND "Battery" AND ("Battery management" OR "Battery State" OR "Battery Performance")) yielded 594 results.

2.5. Selection of Software Tools

After selecting a database, visualization software is selected so as to create networks and maps from the extracted bibliographic data. To carry out bibliometric analysis, CiteSpace and VOSviewer can be used. Citespace is a free-to-use software for the visualization and analysis of literature. It generates structural patterns, maps, and trends in a scientific field. VOSviewer is free software that is used for constructing, analyzing, and visualizing bibliometric maps. The maps can be developed based on citation, co-citation, co-occurrence, or bibliographic coupling basis. The software has a good UI and is easy to use. Figure 4 shows an overview of VOSviewer which was used for this work. The use of VOSviewer was due to the ease of navigation and user-friendly interface.

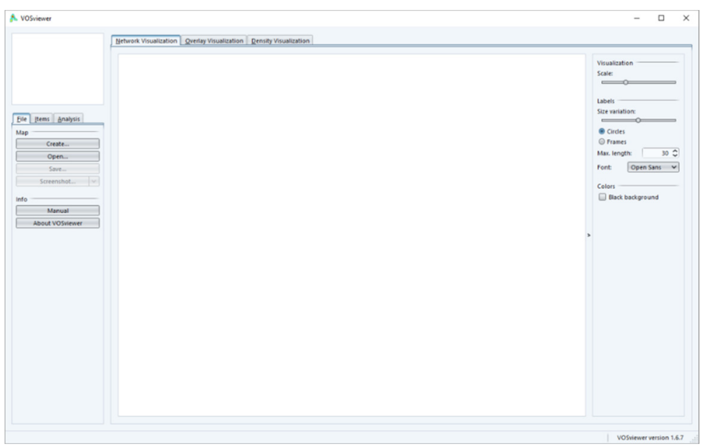

Figure 4: Overview of VOSviewer

2.6. Eligibility Criteria for Selection of Documents

The Scopus database contains a wide range of documents in every field of study. It is therefore important to streamline the results by excluding some literature from the analysis. Table 1 shows the various criteria employed to include and exclude literature.

Table 1: Inclusion and exclusion criteria for selecting Documents

Inclusion criteria	Exclusion criteria
The scope of the paper must be within battery performance analysis using machine learning algorithms	Papers outside this scope were excluded
The papers selected must be written in English	Papers written in other languages were excluded
The papers selected must be journal articles	Review articles were excluded
Papers must be published from 2012 to 2022	Articles published under 2012 are excluded

2.7. Abstracts and Full-text Screening of Selected Articles

To further streamline the results from the database search, abstract and full-text screening is done. The abstracts of the selected projects were carefully screened, and 330 documents were selected which were used for the full-text screening. For the full-text screening, the remaining documents were checked and screened to ensure they correlated to the topic. In the end, 44 documents were excluded, and 286 documents were selected for the visualization.

2.8. Identification and Visualization of Research Gaps

From the 286 documents selected, 191(67%) documents were used to identify gaps, 63(22%) were used to determine research priorities, and 32(11%) were used for both reasons. To visualize the gaps, we make use of VOSviewer to analyze the selected documents. We extract all the keywords contained in the selected article. We set the minimum occurrence of the keywords to be 4. The visualization of the network maps and clusters was obtained. The results are discussed in the next section.

3. RESULTS AND DISCUSSION

3.1. Identification, Prioritization, and Visualization of Research Gaps

Out of a large number of publications available in the machine learning and battery analysis field, 286 were chosen. Keywords extracted from the 286 documents were analyzed to provide network mapping and visualization. A network map of associated keywords is displayed to determine the link strengths of the keywords. Seven clusters are identified in the network and these clusters are carefully analyzed to understand them better. The publication and citation analysis were then analyzed, indicating the most relevant works to this field of study.

3.2. Network Mapping

The mapping of the bibliographic data was carried out using VOSviewer and, a network map was obtained as shown in Figure 5. The most mentioned keywords are “machine learning”, “lithium-ion” and “battery management system” which are central to this field. As seen from the clusters above, publications involving machine learning touches all areas of battery performance.

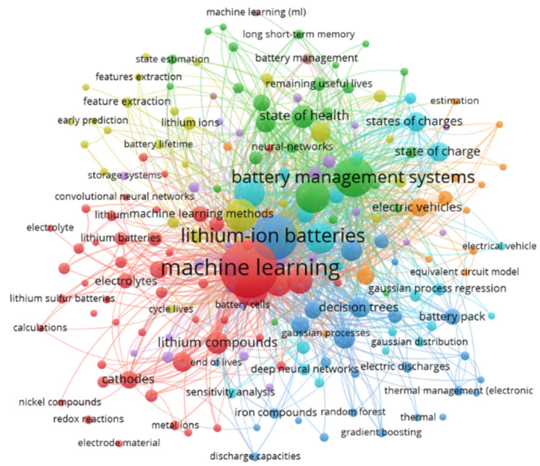

Figure 5: Analysis and mapping of literature related to machine learning for battery performance analysis

3.3. Cluster Analysis

The network mapping of the work resulted in seven clusters of essential topics or areas. Cluster one in Figure 6 shows a strong link between machine learning, battery performance, and lithium-ion, indicating that a lot of research has been done in the area of the battery performance of lithium-ion batteries using machine learning. Cluster two shown in Figure 7 contains topics like battery life cycle, performance, optimization, and cycle lives citing a relationship between the performance and cycle life of a battery. In Figure 8, Cluster three shows a relationship between thermal management and electric discharges. Cluster four shown in Figure 9 contains topics that show a strong link between data-driven methods and machine learning models to battery health. Cluster five shown in Figure 10 highlights the relationship between neural networks and battery states estimation and prediction. In cluster six, shown in Figure 11 there is a strong link between battery management systems, electrical vehicles, and machine models such as support vector machines. A connection between the reliable operation of the battery and the estimation of the state of charge using the Gaussian process is shown in cluster seven depicted in Figure 12. The mostly used machine models are support vector machines, neural networks, Gaussian progress regression, deep neural networks, and long short-term memory.

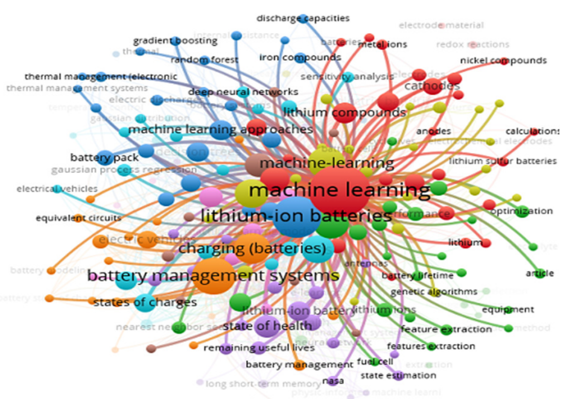

Figure 6: Cluster one

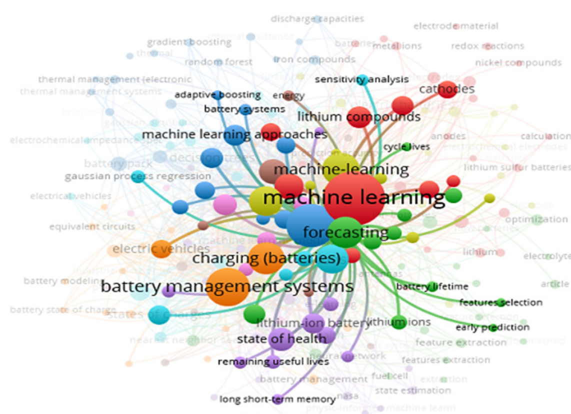

Figure 7: Cluster two

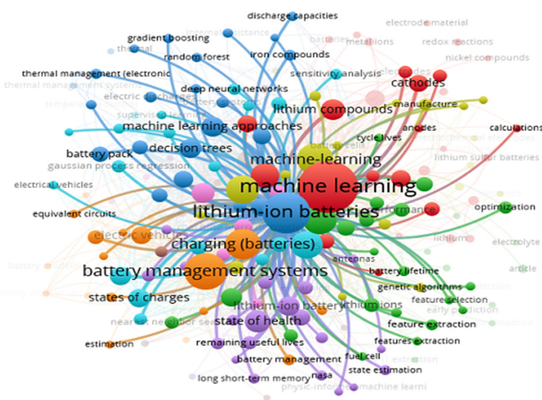

Figure 8: Cluster three

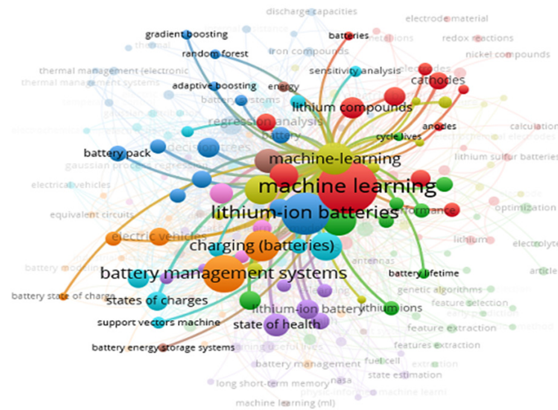

Figure 9: Cluster four

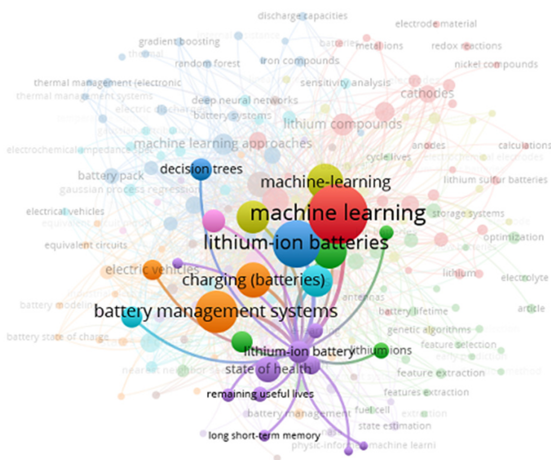

Figure 10: Cluster five

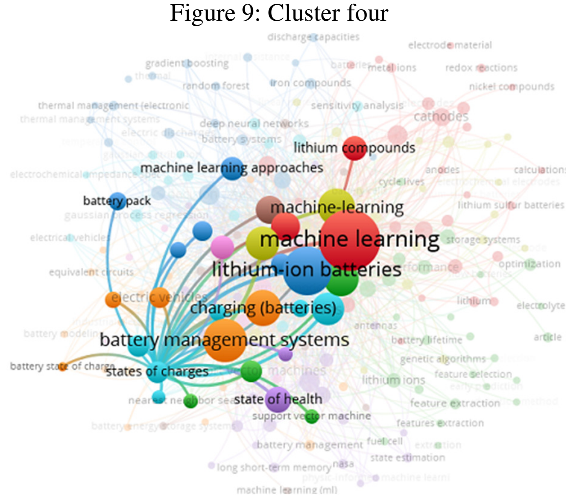

Figure 11: Cluster six

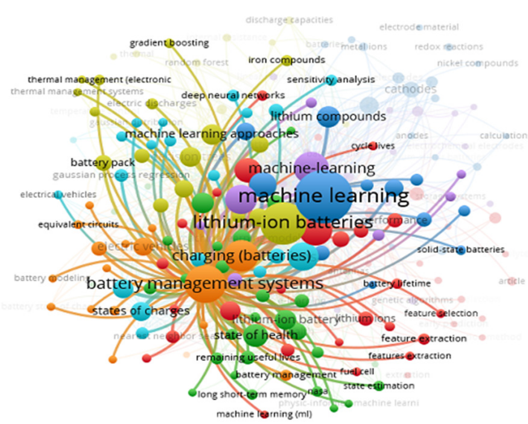

Figure 12: Cluster seven

3.4. Publication Analysis

The number of publications in the field of machine learning for battery performance analysis has increased over the years as this field has been gaining wide interest in recent years. Figure 13 shows the number of publications

A. Obayuwana and J.U. Okoye / Nigerian Research Journal of Engineering and Environmental Sciences
8(1) 2023 pp. 92-101

per year since 2013. Up till 2013, there were less than ten publications identified which indicates that the field of study is a recent one that is still open to more developments and innovations. The number of publications has increased for every consecutive year, with that trend expected to continue in the coming years since there is a demand for cleaner and renewable sources of energy. Being a research hotspot, a lot of authors have published works in this field. Zhang Y. leads the way with the highest number of publications. Figure 14 shows the ten authors with the greatest number of works in this research.

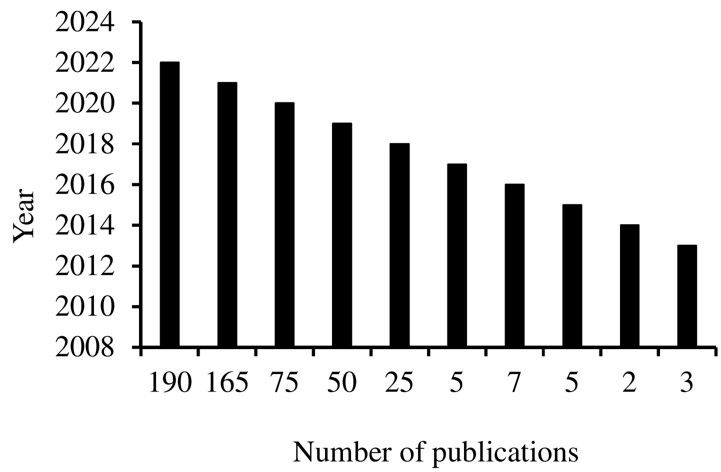

Figure 13: Number of publications per year

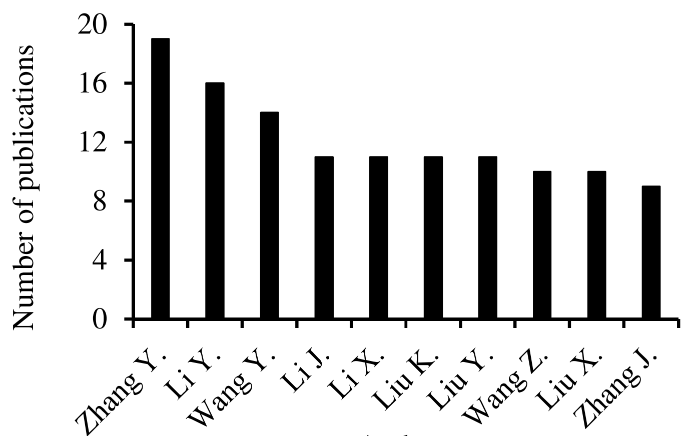

Figure 14: Top ten authors according to the number of works published

3.5. Citation Analysis

The impact or quality of literature is accessed by considering the number of citations for that publication. It shows the importance and relevance of the work in that field of study. The work by Nuhic et al. (2013) is the most cited work in this field with about 413 citations. Table 2 shows the top 10 cited works in this field. The findings indicate a large amount of research work done on machine learning, battery management, and lithium-ion batteries. The state of health and state of charge estimation of a battery using machine learning are well represented. The number of publications indicates a yearly increase as the field of study becomes more widely explored. The map of the bibliographic data shows that research gaps exist in the area of performance analysis of lead-acid batteries. In battery usage, lithium-ion batteries are more widely used due to their long lifespan, fast charging rate, and higher efficiency. The limited availability of public datasets also accounts for this. Research gaps also exist in the area of sensitivity analysis, and the use of cycle life for battery state prediction

Table 2: Top 10 publications with the highest citations

S/No.	No. of citations	Title
1.	413	Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods
2.	334	Battery health prognosis for electric vehicles using sample entropy and sparse Bayesian predictive modeling
3.	318	Long Short-Term Memory Networks for Accurate State-of-Charge Estimation of Li-ion Batteries
4.	289	State-of-charge estimation of Li-ion batteries using deep neural networks: A machine learning approach
5.	284	A novel multistage Support Vector Machine based approach for Li ion battery remaining useful life estimation
6.	216	Advanced Machine Learning Approach for Lithium-Ion Battery State Estimation in Electric Vehicles
7.	204	Random forest regression for online capacity estimation of lithium-ion batteries
8.	197	A Data-Driven Approach with Uncertainty Quantification for Predicting Future Capacities and Remaining Useful Life of Lithium-ion Battery
9.	181	Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning
10.	170	Modified Gaussian Process Regression Models for Cyclic Capacity Prediction of Lithium-Ion Batteries

4. CONCLUSION

This paper carried out bibliometric analysis in the field of machine learning for better performance analysis. Related literature was retrieved from the Scopus database. The search text used was "Machine Learning", "Battery performance", and "Battery management system". A large number of results were filtered and carefully selected, and the keywords were extracted from the selected articles. Network mapping and visualization of keywords were done using the VOSviewer. This visualization revealed a strong relationship between machine learning, battery management systems, and lithium-ion. Lead-acid batteries, sensitivity analysis, and cycle life are identified as the research gaps present in battery performance analysis using machine learning. It is expected that more research will be done in these areas in the coming years to improve the performance and capability of batteries.

5. CONFLICT OF INTEREST

There is no conflict of interest associated with this work.

REFERENCES

Boutaba, R., Salahuddin, M.A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F. and Caicedo, O.M. (2018). A comprehensive survey on machine learning for networking: evolution, applications, and research opportunities. *Journal of Internet Services and Applications*, 9(1), pp. 1-99.

Chen, C., Zuo, Y., Ye, W., Li, X., Deng, Z. and Ong, S.P. (2020). A critical review of machine learning of energy materials. *Advanced Energy Materials*, *10*(8), p. 1903242.

Dunjko, V. and Briegel, H.J. (2018). Machine learning & artificial intelligence in the quantum domain: a review of recent progress. *Reports on Progress in Physics*, *81*(7), p. 074001.

Flah, M., Nunez, I., Ben Chaabene, W. and Nehdi, M.L. (2021). Machine learning algorithms in civil structural health monitoring: a systematic review. *Archives of computational methods in engineering*, *28*, pp. 2621-2643.

Goh, G.B., Hodas, N.O. and Vishnu, A. (2017). Deep learning for computational chemistry. *Journal of computational chemistry*, *38*(16), pp. 1291-1307.

Issame, M. (2022). *Object detection for quadrotor using deep learning.* Masters thesis, University of M'sila.

Khumprom, P. and Yodo, N. (2019). A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm. *Energies*, *12*(4), p.660.

Laayouj, N., Jamouli, H. and El Hail, M. (2016). New prognostic framework for degradation assessment and remaining useful life estimation of photovoltaic module. In: *24th Mediterranean Conference on Control and Automation (MED)* (pp. 378-383). IEEE.

Lavecchia, A. (2019). Deep learning in drug discovery: opportunities, challenges and future prospects. *Drug discovery today*, *24*(10), pp. 2017-2032.

Morgan, D. and Jacobs, R. (2020). Opportunities and challenges for machine learning in materials science. *Annual Review of Materials Research*, *50*, pp. 71-103.

Obayuwana, A. and Monica, G.O., 2021. A Comparative Study of Machine Learning Models for Solar Power Generation Forecasting Using Weather Parameters: A Case of Benin City. *NIPES Journal of Science and Technology Research*, *3*(4), pp. 141-152.

Sindhu, V., Nivedha, S. and Prakash M. (2020). An Empirical Science Research on Bioinformatics in Machine Learning. *Journal of Mechanics of Continua and Mathematical Sciences*, 7, pp. 86-94.

Wang, Z.H., Hendrick, Horng, G.J., Wu, H.T. and Jong, G.J. (2020). A prediction method for voltage and lifetime of lead–acid battery by using machine learning. *Energy Exploration & Exploitation*, *38*(1), pp. 310-329.