5-20-2021

A Bibliometric Survey on the Use of Long Short-Term Memory Networks for Multivariate Time series forecasting

Vidur Sood Mr.
Symbiosis Institute of Technology, Symbiosis International (Deemed University),
vidur.sood.btech2018@sitpune.edu.in

Manobhav Mehta Mr.
Symbiosis Institute of Technology, Symbiosis International (Deemed University),
manobhav.mehta.btech2018@sitpune.edu.in

Vedansh Mishra Mr.
Symbiosis Institute of Technology, Symbiosis International (Deemed University),
vedansh.mishra.btech2018@sitpune.edu.in

Akash Upadhyay Mr.
Symbiosis Institute of Technology, Symbiosis International (Deemed University),
upadhyay.akash.btech2018@sitpune.edu.in

Shilpa Hudnurkar
Symbiosis Institute of Technology, Symbiosis International (Deemed University),
shilpa.hudnurkar@sitpune.edu.in

Authors

Vidur Sood Mr., Manobhav Mehta Mr., Vedansh Mishra Mr., Akash Upadhyay Mr., Shilpa Hudnurkar, Shilpa Gite Dr., and Neela Rayavarapu Dr.

A Bibliometric Survey on the Use of Long Short-Term Memory Networks for Multivariate Time series forecasting

Vidur Sood
Department of Electronics and Telecommunication, Symbiosis Institute of Technology, Symbiosis International (Deemed University), (SIU) Lavale, Pune, Maharashtra, India vidur.sood.btech2018@sitpune.edu.in

Manobhav Mehta
Department of Electronics and Telecommunication, Symbiosis Institute of Technology, Symbiosis International (Deemed University), (SIU) Lavale, Pune, Maharashtra, India manobhav.mehta.btech2018@sitpune.edu.in

Vedansh Mishra
Department of Electronics and Telecommunication, Symbiosis Institute of Technology, Symbiosis International (Deemed University), (SIU) Lavale, Pune, Maharashtra, India vedansh.mishra.btech2018@sitpune.edu.in

Akash Upadhyay
Department of Electronics and Telecommunication, Symbiosis Institute of Technology, Symbiosis International (Deemed University), (SIU) Lavale, Pune, Maharashtra, India upadhyay.akash.btech2018@sitpune.edu.in

Shilpa Hudnurkar*
Department of Electronics and Telecommunication, Symbiosis Institute of Technology, Symbiosis International (Deemed University), (SIU) Lavale, Pune, Maharashtra, India shilpa.hudnurkar@sitpune.edu.in

Shilpa Gite
Department of Computer Science, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Faculty, Symbiosis Centre for Applied Artificial Intelligence (SCAAI), (SIU) Lavale, Pune, Maharashtra, India
shilpa.gite@sitpune.edu.in

Neela Rayavarapu
Ph.D. Supervisor, Symbiosis International (Deemed University), (SIU) Lavale, Pune, Maharashtra, India neela.raya27@gmail.com

Abstract

In this paper, we aim to review and analyze the publications related to the utilization of Long Short-Term Memory (LSTM) networks for multivariate time series forecasting. The purpose of this bibliometric survey was to study how technology in the field of LSTM has evolved over the years. There were 242 research papers published, by over 50 researchers, over 6 years, on the topic of *"Multivariate time series forecasting using LSTM"*. The majority of these papers were published between the years 2018 and 2020. The Scopus database was utilized for analyzing recent trends in this area and to determine the model that would be best suited for weather forecasting applications. Through this study, we aim to shortlist various models that have shown consistent reliability and accuracy while utilizing multivariate time series data for prediction. These models can then be employed for other forecasting applications.

Methods: The articles on the subject were obtained from one of the most popular databases i.e., Scopus. The Scopus search resulted in publications from the years 2015 to 2021. All these years were considered for the analysis. The Scopus analyzer is used for the analysis of the database based on different classes such as source, subject, country, etc. The analysis based on co-authorship, co-occurrences, and citation analysis was achieved by employing VOS viewer Version 1.6.16.

Results: In the study, publications on Multivariate time series forecasting using LSTM, obtained from the Scopus database, were between the years 2015 and 2021. A year-wise increasing trend in the number of publications was observed. The statistical analysis and network analysis showed that the highest number of publications and the largest variety of papers were published in the year 2020. China was the highest contributor followed by the USA and India.

Conclusion: The outcome of the Scopus database was 242 articles with most of the publications in the English Language. The statistical analysis was done for authors, documents, countries, affiliations, funding sponsors. The network analysis indicated the interconnections between completely different parameters such as authors and sources.

1. Introduction

A Multivariate time series, as the name proposes, is a series with more than one time-subordinate variable (K. Wang et al., 2019). Each variable relies upon its past qualities and may likewise have some reliance on different factors (Amin et al., 2019). This reliance is utilized for determining future qualities. Time series forecasting is a challenging task as it has trend, periodicity, randomness, or cyclicity (Munkhdalai et al., 2019). Sometimes, the time series consists of all these components and in this case the decomposition of the time series is required. If there are multiple variables in the time series, the complexity increases further and, along with decomposition, other aspects also need to be explored (Y. S. Wang et al., 2020). There are various techniques for decomposing a time series. With the recent advances in Artificial Intelligence, many supervised, unsupervised, and reinforcement learning algorithms have been developed (Wojtkiewicz et al., 2019). Some of these algorithms have been found suitable for time series forecasting. One such technique is the Recurrent Neural Network (RNN).

One of the special kinds of RNN is Long Short-Term Memory (LSTM) Network. It is capable of learning future dependencies in data (Wu, 2020). This is possible because the recurring module of the model is a combination of four layers interacting with each other (Yu et al., 2019). Because of the long-term memory, it finds applications in sequence-to-sequence learning, time series prediction, etc.

Several researchers have worked in the area of multivariate time series forecasting using LSTM networks. The use of LSTM has been found in many applications such as in making weather predictions for small areas (Li., 2020), for cloud resource forecasting to effectively manage cloud resources (Tran et al., 2018), wind turbine and grid interaction forecasting (Y. Wang et al., 2018), urban particulate matter forecasting (Li et al., 2020), short term load forecasting (Santra & Lin, 2019), and short term solar photovoltaic power plant's power prediction (Chen et al., 2020). It was observed from the analysis, that the application of multivariate time series using LSTM to style weather prediction systems could be a significant research topic within the area of research (Liu et al., 2020) especially in countries such as India, the United States, Europe, and Canada (Alhirmizy & Qader, 2019). There has been a significant increase in the number of research studies conducted on LSTM (Zhu et al., 2017).

A bibliometric approach is presented in this paper to investigate efforts made by researchers to analyze the multivariate time series forecasting using LSTM networks. The goal of this study was to analyze the publications, to date, on multivariate time series (MTS)

forecasting using LSTM networks for various applications. This would facilitate other researchers in applying LSTM to their respective domains. Mentioned below are some of the useful features of bibliometric analysis.

- For elaborating upon the different types of documents in academic publications within the exploration field.
- To know the extent of languages used for research publications in a particular domain.
- Examine the pattern in which the research papers were published over the period.
- Investigation upheld geographical areas involves the nations of the reality where the examination has been done before.
- Analysts who made contributions to the area of examination.
- Models of the distributions depending on the association or affiliations
- Papers that are referred to as the absolute best on the topic.

This paper provides a bibliometric survey on " Multivariate time series using LSTM" to understand the areas where LSTM networks have been employed. The next section shows the information gathered from the Scopus data set and the evaluation performed on the data set.

2. Materials and Methods

2.1 Search Procedure

Scopus, JSTOR, Web of Science, and Google Scholar are some of the popular databases worldwide. We have used the largest database, Scopus from the above-mentioned databases. Over 200 publication results have been obtained using the keywords for this search. For searching databases across the world, different keywords were used. This information was employed for the analysis. The keywords were segregated further into 2 types i.e primary keywords and secondary keywords.

Table 1. Details of Keywords

Primary Keyword	Multivariate Time-Series
Primary Keyword (Using the Operator)	(AND) LSTM
Secondary Keyword	Prediction

Source: The Scopus database (accessed on 28th April 2021)

2.2 Search Outcomes

A total of 242 documents were obtained from the Scopus database based on the primary and secondary keywords mentioned in the above section. The publications were found to be from the year 2015 to 2021. The following tables and figures show the analysis of search results based on different parameters. Table 2 presents the number of publications in different languages. It can be observed that the maximum publications are in the English language followed by Chinese. Portuguese and other languages which have one publication each.

Table 2. Language of Publication

Publishing Language	Count of Publications
English	138
Chinese	6
Portuguese	1
Others	97

Source: The Scopus database (accessed on 28th April 2021)

3. Analysis of Performance

For further analysis, the Scopus database was accessed on 28th April 2021 and the results obtained on the basis of the keywords mentioned in Table 1 were used for carrying out the statistical and network analysis. The product that was utilized for the information base investigation, further to the examination from Scopus, was the VOS viewer 1.6. The network of co-authors, co-occurrences, co-citations, bibliometric couplings, etc. are often effectively analyzed by VOS viewer 1.6. Different forms of analysis were performed.

➢ Statistical Analysis of Database

- Analysis by source
- Analysis by subject area
- Analysis per year
- Analysis on the basis of type of documents
- Analysis by country
- Analysis based on number of documents by author
- Analysis by affiliations
- Analysis by sponsorship for the research

- **Network Analysis of Database**
  - Analysis of co-authorship
  - Analysis of co-occurrences
  - Analysis of citations
  - Analysis of co-citations

4. Results

4.1. Statistical Analysis

Two different techniques were used for the analysis i.e., statistical analysis and network analysis of the databases.

4.1.1 Analysis by Source

Figure 1 below shows year-wise publication statistics from multiple sources like IEEE access, ACM International Conference proceeding series, Communications in Computer, and Information Science. etc. A graphical representation can be seen in Figure 1 with the number of documents published per year.

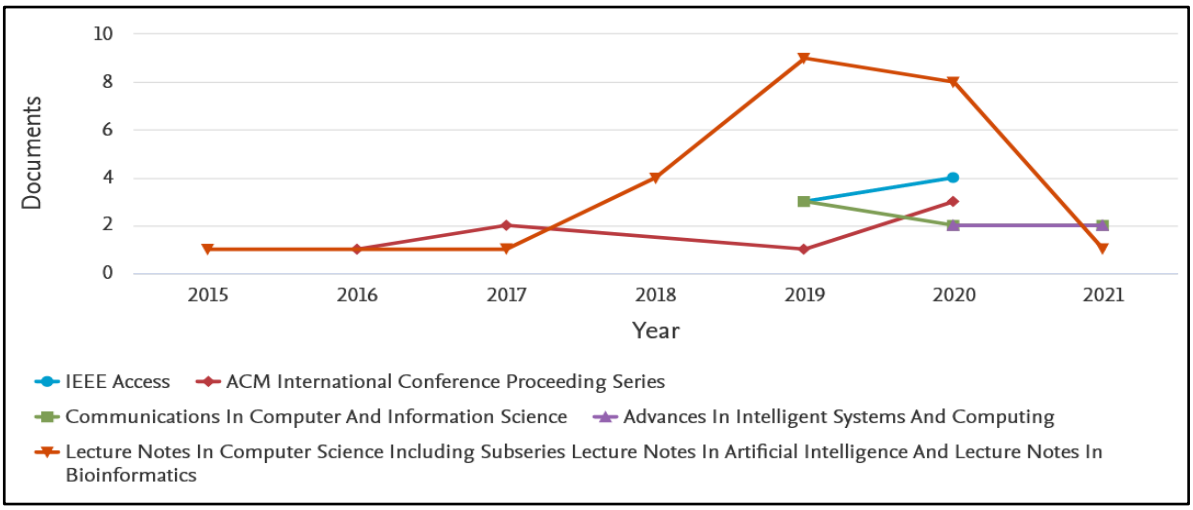

Figure 1. Analysis of publications per year on MTS forecasting using LSTM

4.1.2 Analysis by Subject Area

On LSTM technology, maximum papers were found under computer science (35.4%), 18.4% of the papers were in Engineering and the remaining documents were published in other subject areas. The main reason for this is often that the subject is most related to the sector of engineering, computing, and arithmetic.

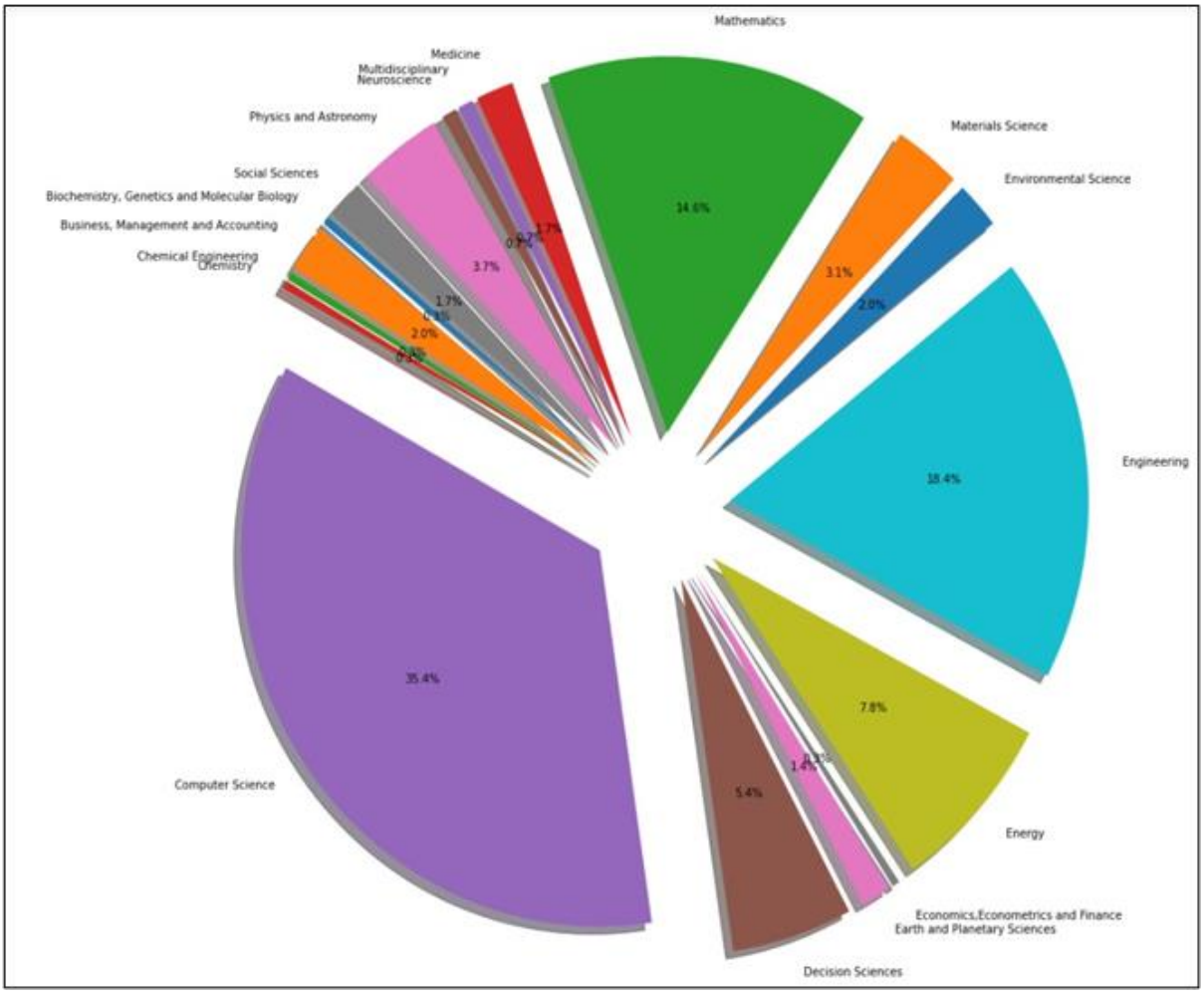

Figure 2: Analysis of documents on MTS forecasting using LSTM

4.1.3 Analysis by year

Documents were collected from the Scopus database including different sources like conferences, journals, research papers, etc. from the year 2015 to 2021. The Table 3 below gives the statistical information, and the graphical representation is shown in Figure 3. The highest number of publications was in the year 2020 as observed from the analysis.

Table 3: Year of Publication

Year	Number of Publications
2015	3
2016	2
2017	9
2018	17
2019	78
2020	98
2021	35

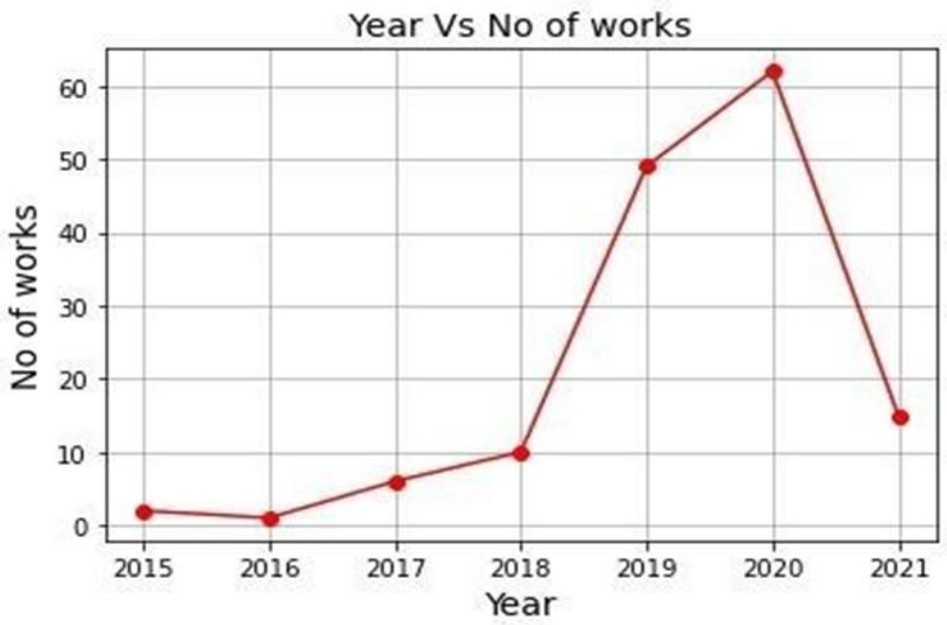

Figure 3: Analysis of documents by year published on MTS forecasting using LSTM

4.1.4 Analysis on the basis of the Type of Documents

The majority of the contributions were from conference papers, followed by the articles, as observed from the analysis.

Table 4: Document Analysis by Type

No.	Type of Document	Publications
1.	Conference Papers	131
2.	Articles	100
3.	Conference Reviews	9
4.	Book Chapters	1
5.	Reviews	1

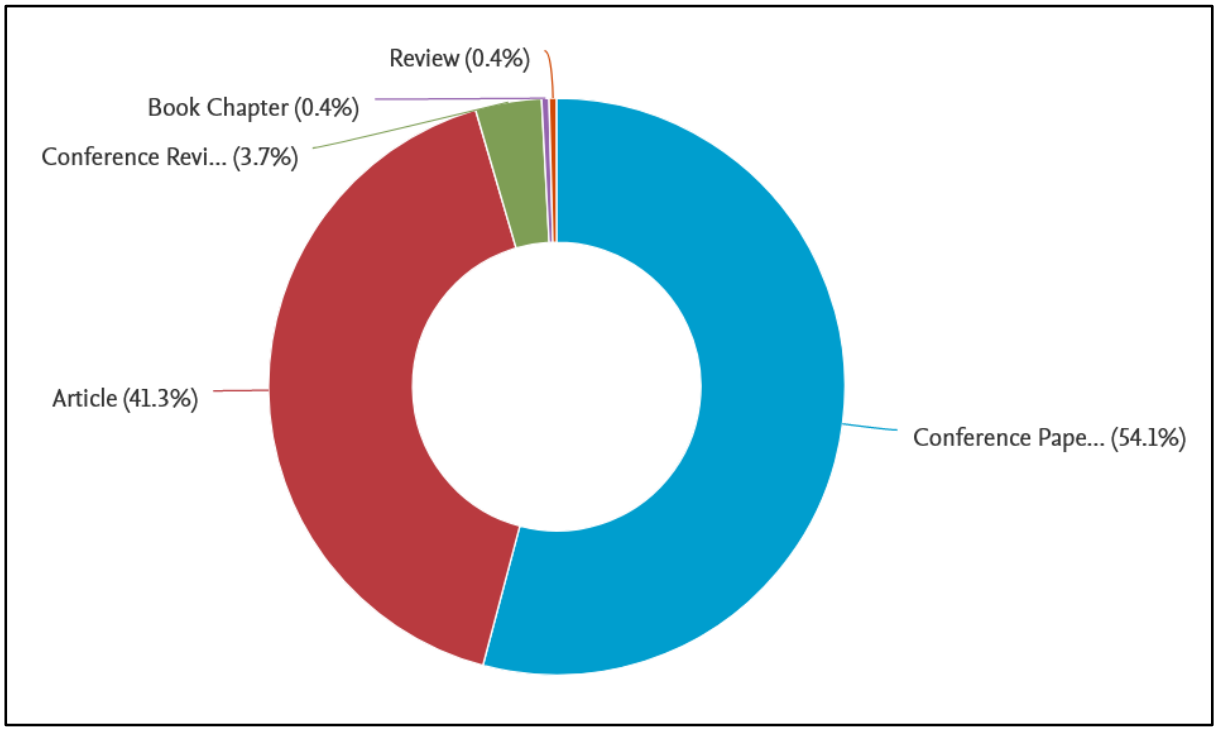

Figure 4: Analysis of documents by type on MTS forecasting using LSTM

4.1.5 Analysis by Country

By considering the number of publications distributed, this Scopus data set was investigated for the different countries from where these papers originated. For the chosen timeline it was observed that a good number of documents were published by China, followed by the United States and India (Liu., 2019).

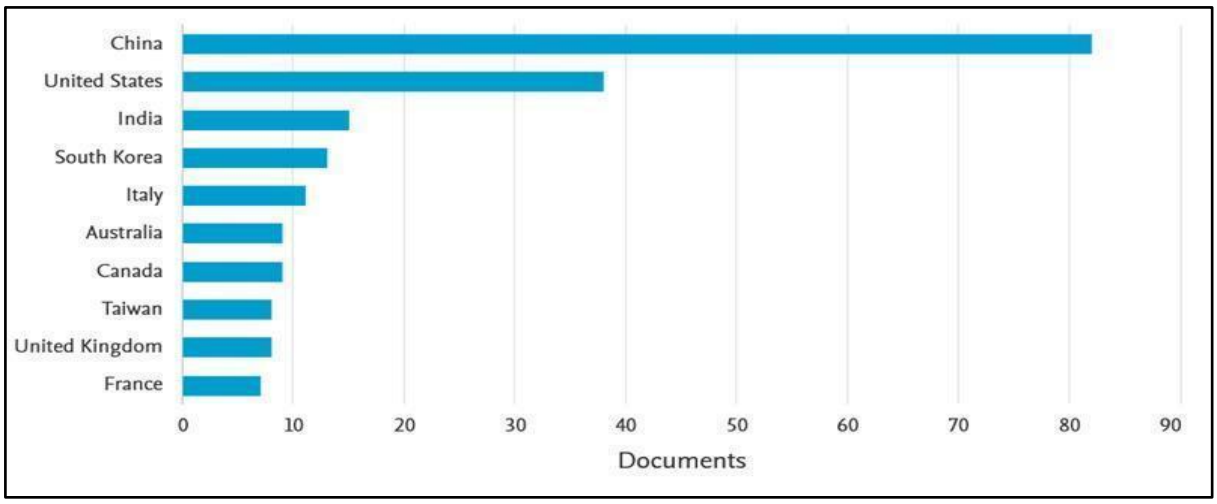

Figure 5: Analysis of documents by country on MTS forecasting using LSTM by Sources

4.1.6 Analysis based on Number of Documents by Author

The top 10 authors were identified as shown in Figure 6 below. Each of these 10 authors has the same number of publications as shown in the graph.

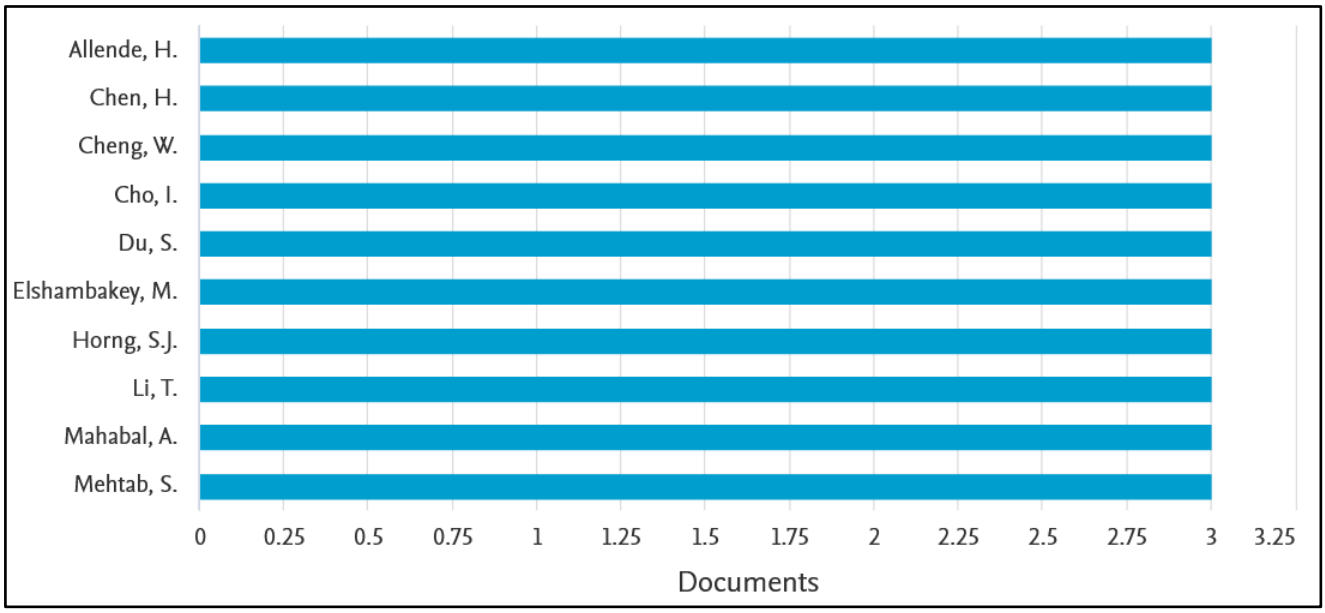

Figure 6: Analysis of documents by author on MVTS forecasting using LSTM by Sources.

4.1.7 Analysis by Affiliations

The prime affiliations were considered for this analysis. It was found that the 'Southwest Jiaotong University' and 'Chinese Academy of Sciences' had done most of the work done in this field. Together they were affiliated with 11 publications. The Ministry Education of China, Beijing University and South China University of Technology had also contributed significantly to study in this field.

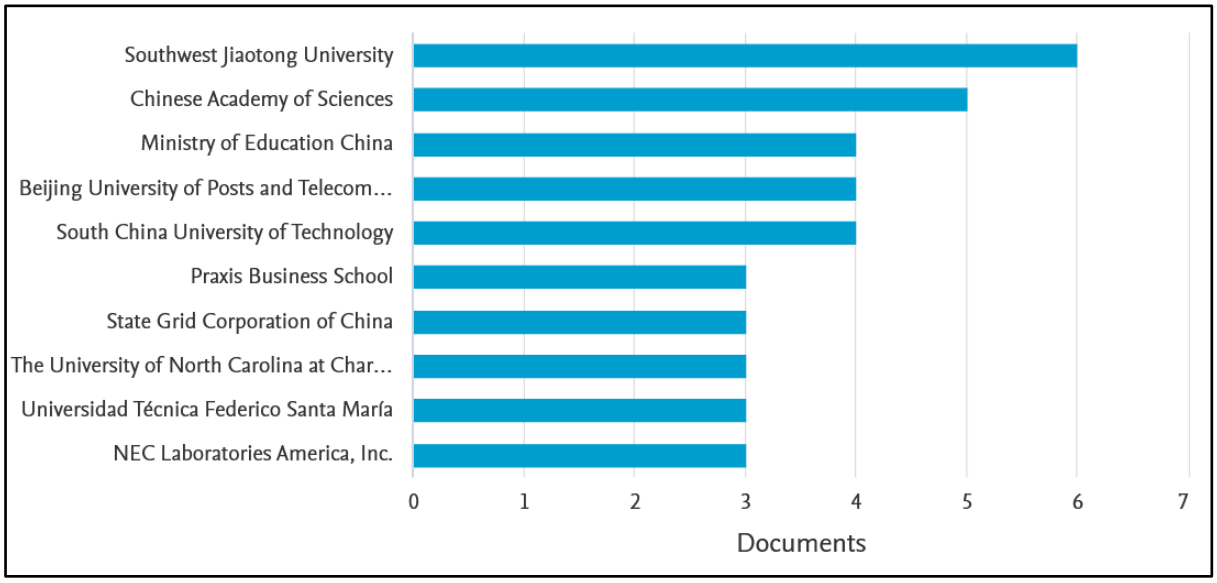

Figure 7: Analysis of documents by affiliation on MTS forecasting using LSTM

4.1.8 Analysis by Sponsorship for the Research

China has the highest funding from the National Natural Science Foundation of China. From the analysis, it was found that the majority of the funded papers were from research institutes. The Ministry of Education of the People's Republic of China and Ministry of Science and Technology of the People's Republic of China are funded equally by China.

Table 5: Document Analysis by Funding Sponsor

Name of Institute	Number of projects funded
National Natural Science Foundation of China	41
Ministry of Education of The People's Republic of China	11
Ministry of Science and Technology of The People's Republic of China	11
National Science Foundation	10
Fundamental Research Funds for The Central Universities	9
European Commission	8
Ministry of Finance	8
Name of Institute	**Number of projects funded**
National Key Research and Development Program of China	7
Horizon 2020 Framework Programme	6
Ministry of Science and ICT, South Korea	4

Ministry of Science and Technology, Taiwan	4
Ministry of Science, ICT And Future Planning	4
National Basic Research Program of China (973 Program)	4
National Research Foundation of Korea	4
State Grid Corporation of China	4

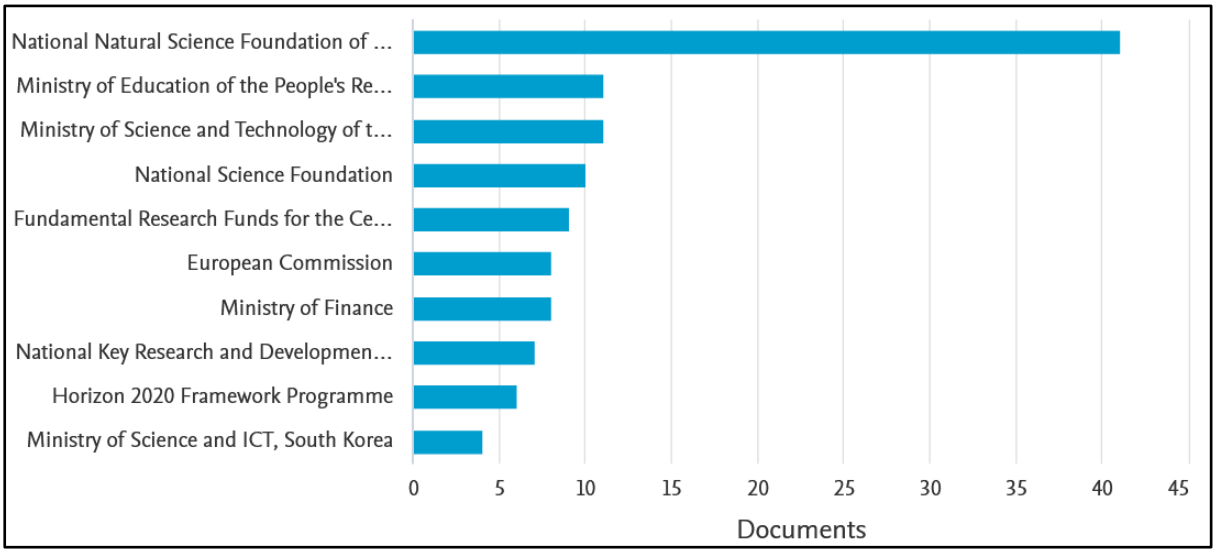

Figure 8: Analysis of based on institutes funded for research on MTS forecasting using LSTM

4.2. Network Analysis

4.2.1 Network Analysis of Co-Authorship by Authors

When considering authors, organizations, and countries analysis as the basis for analysis, a document that had too many authors (25 authors in one case) was ignored. VOS viewer allows the user to determine the threshold value case-sensitively. An author with a minimum of 3 documents was considered as the threshold value in this case. It was observed that, within the total of 751 authors, only 32 authors met these threshold criteria. It has the most noteworthy number of publications equal to 9 in this analysis.

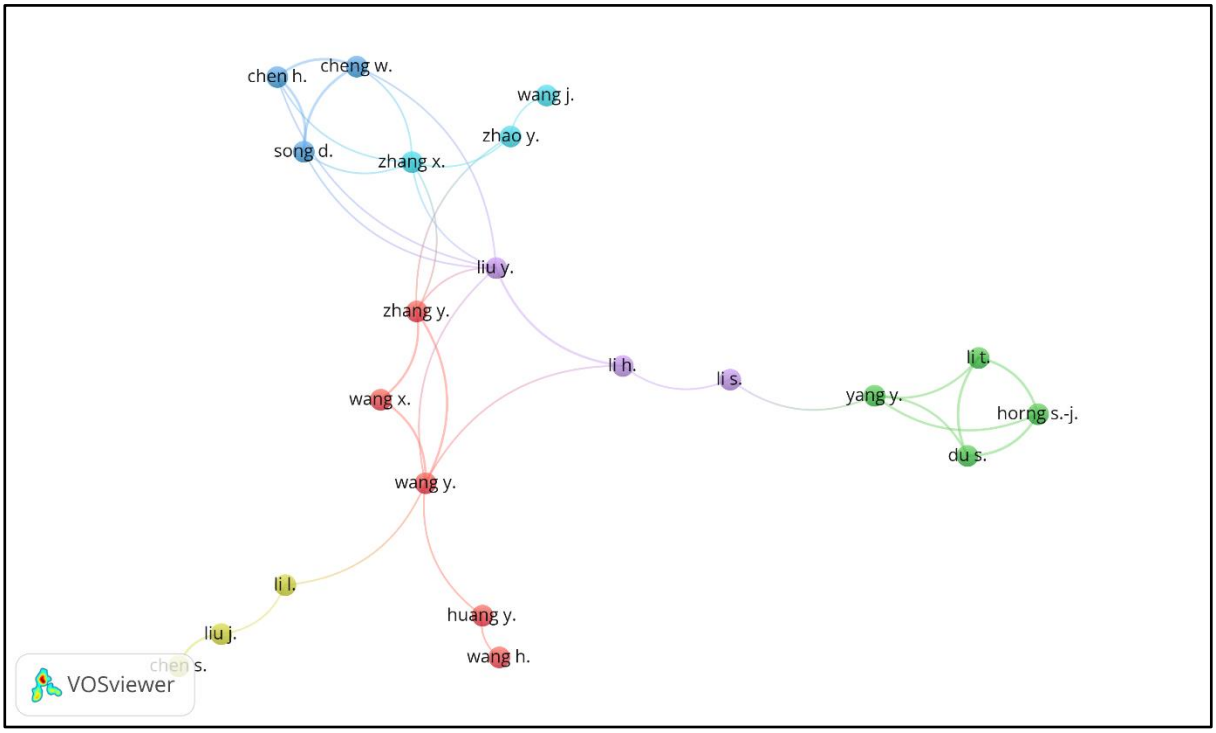

Figure 9: Network Analysis by Co-Authorship

4.2.2 Analysis of Co-Occurrences

A. Co-occurrence by All Keywords

Keywords are the most important feature of any article. Co-occurrences of different keywords were analyzed. 12 was the threshold considered here. It was found that 36 keywords met the threshold out of 1993 keywords considered.

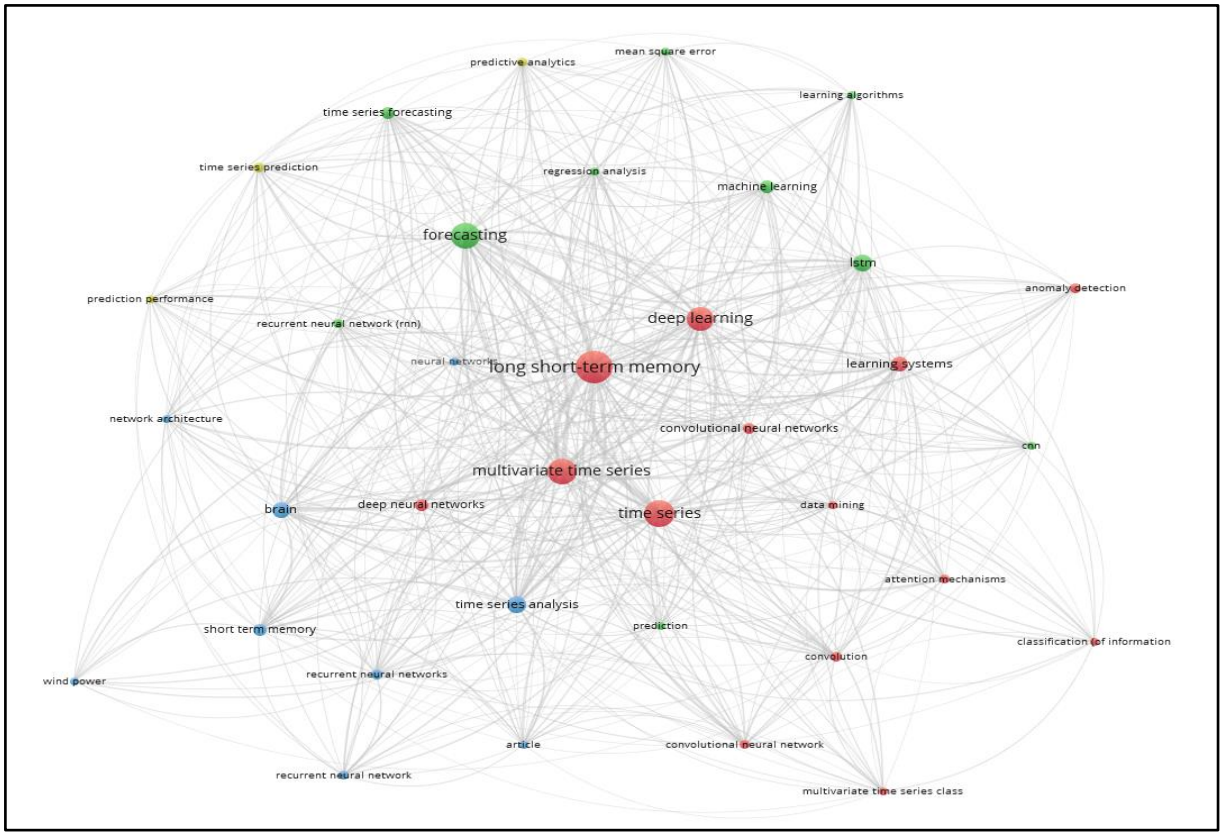

Figure 10: Network Analysis of Co-occurrence for All Keywords

B. Co-occurrence by Author Keywords

For this Co-occurrence, the base limit was 4 for each author. Of the 557 keywords, the limit was exceeded by 35 keywords.

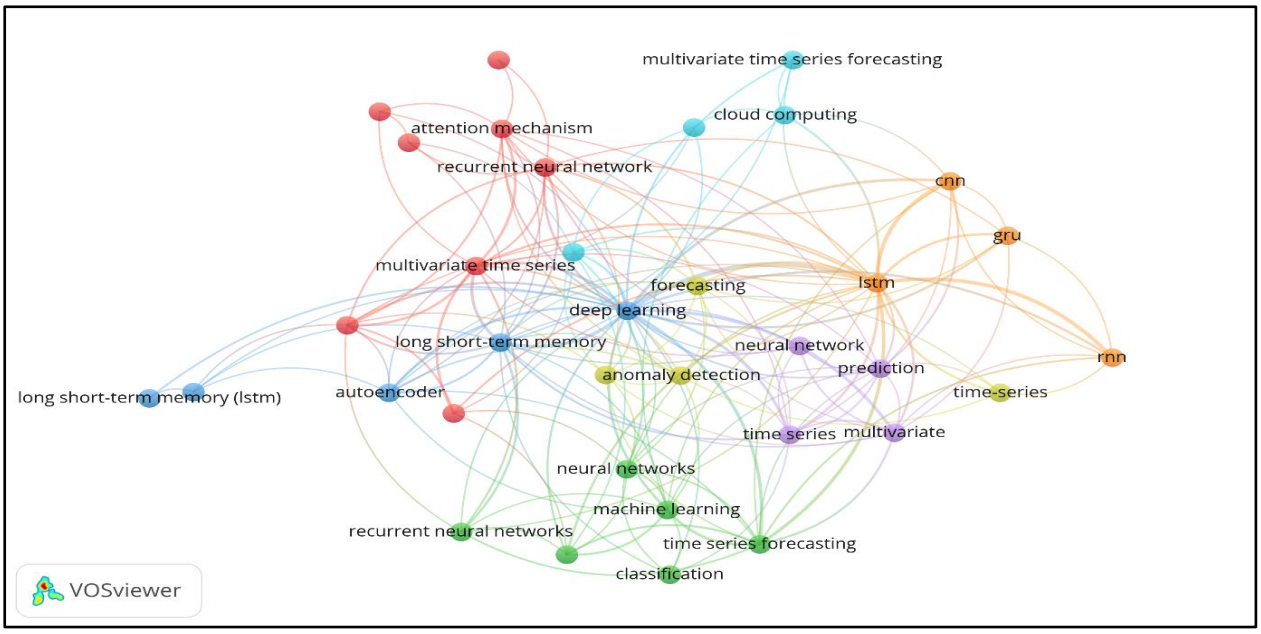

Figure 11: Network analysis (CO-Occurrence for Author Keywords)

C. Co-Occurrence by Index Keywords

For this analysis, the threshold set was 11. Index keywords were 1623 in total. Co-concurrence in terms of these words outcomes 36 keywords that met the threshold.

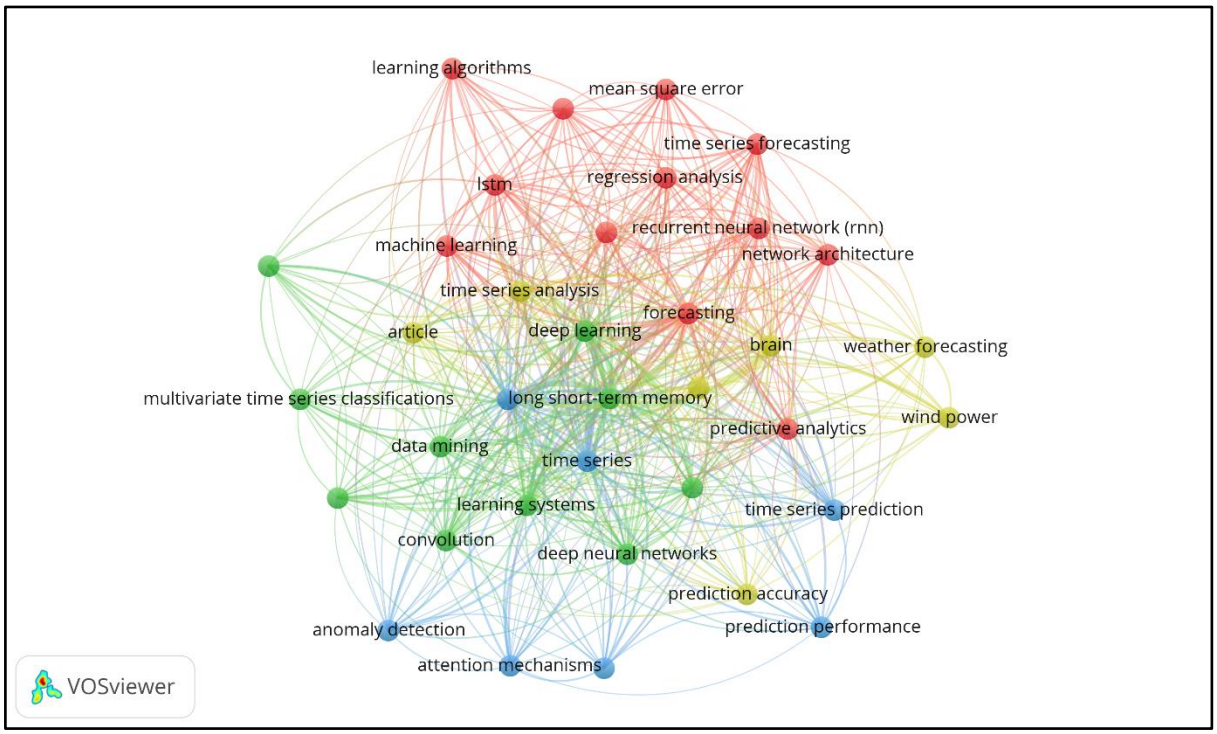

Figure 12: Network Analysis (Co-Occurrence for Index Keywords)

4.2.3 Analysis of Citations.

A. Citation by Documents

The network analysis was done for citations of documents and sources. There were a total of 242 documents. 9 citations was taken as the threshold and only 36 documents met the threshold.

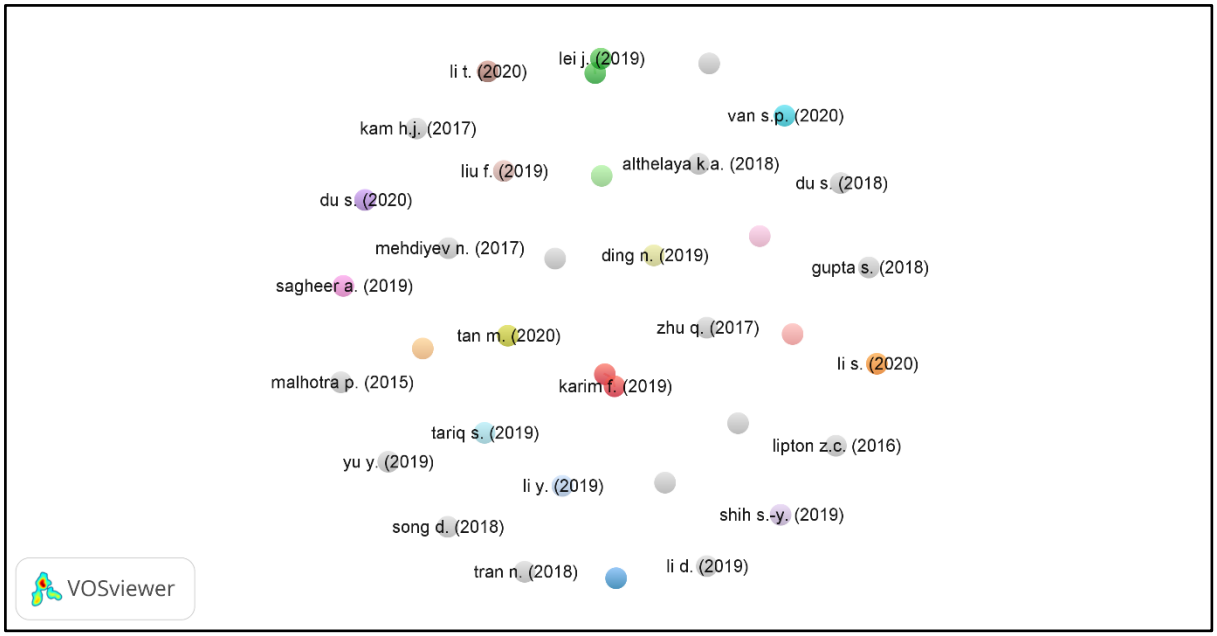

Figure 13: Network Analysis (Citation by documents)

B. Citation by Sources

Reference examination of sources was acquired by considering the limit of 3 references for every source. Out of these 166 sources, only 13 achieved the set limit.

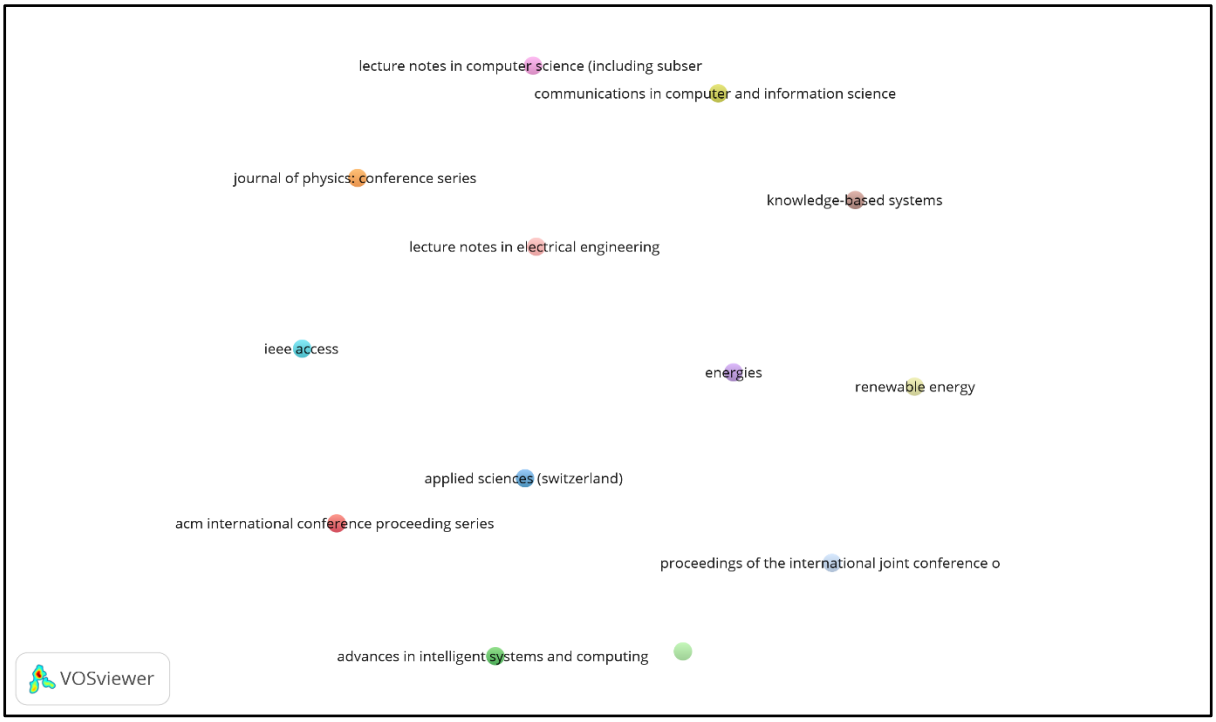

Figure 14: Network Analysis(Citation by Sources)

4.2.4 Analysis of bibliographic coupling

A. Bibliographic coupling of documents

In this analysis, 36 sources met the threshold amongst a total of 242 documents. The threshold considered here was 9 documents.

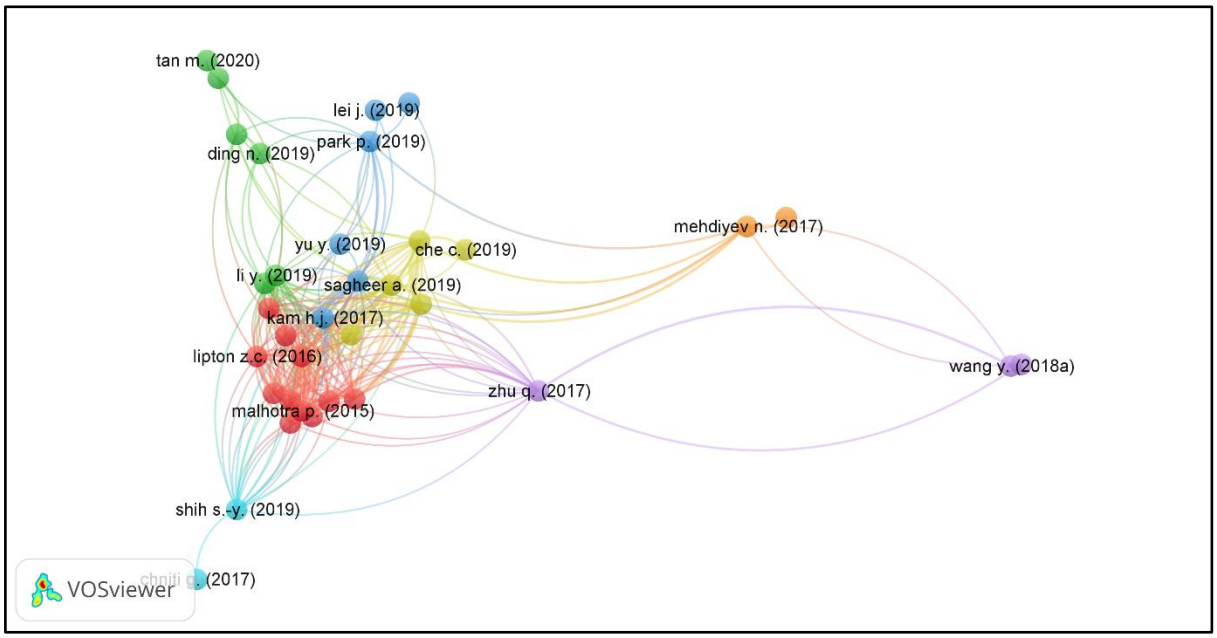

Figure 15: Network Analysis (Bibliographic Coupling Of Document)

B. Bibliographic Coupling of Sources

In this analysis, 27 sources met the limit among some of 166 sources. The threshold considered here is 2 documents per source.

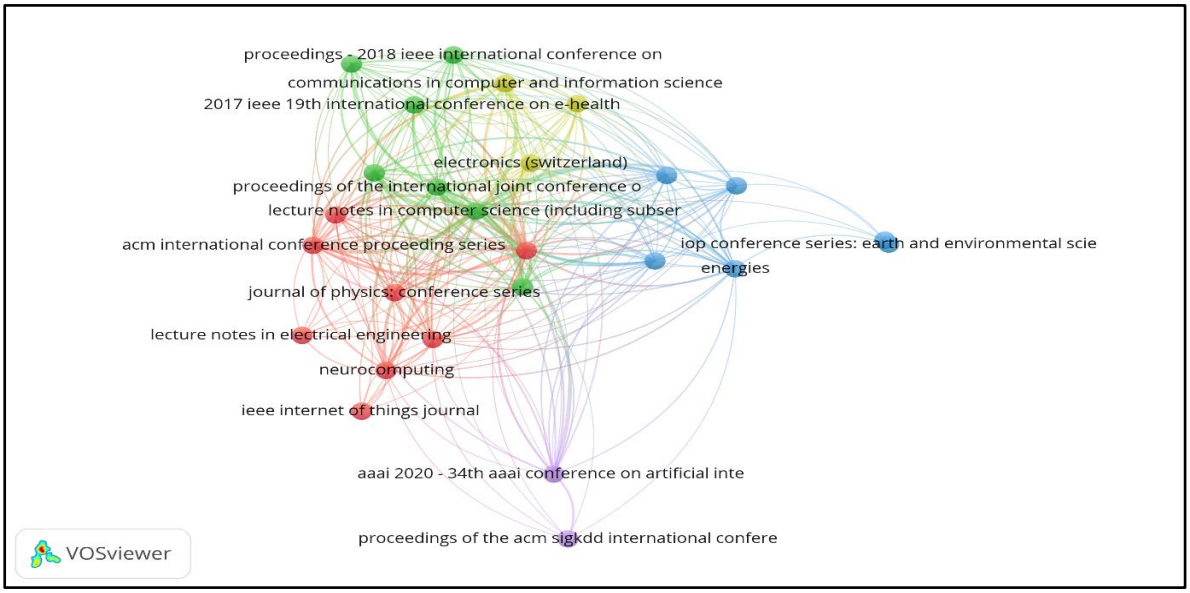

Figure 16: Network analysis showing coupling of Sources

4.2.5 Analysis of Co-Citation

A. Co-citation of Cited References

In this database, there were a total of 6696 cited references. By keeping the threshold of a minimum of 4 citations per cited references, a total of 22 achieved the limit.

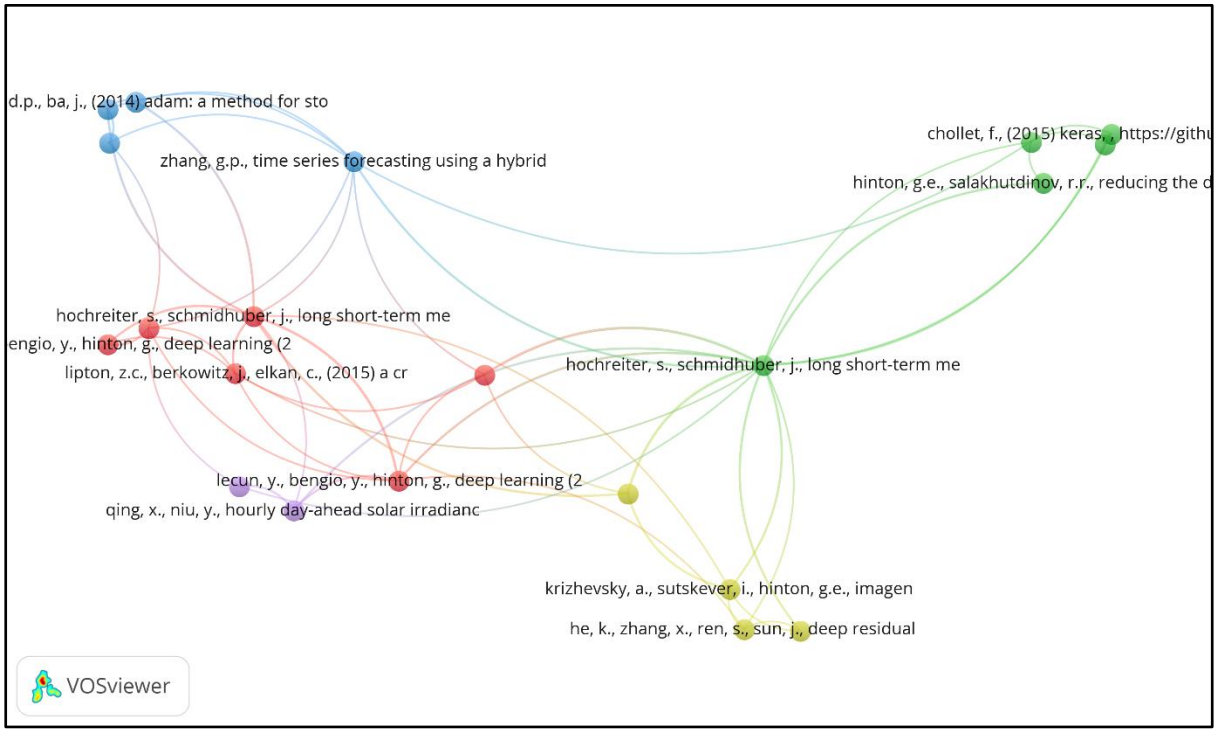

Figure 17: Network Analysis (Co-citation of cited references)

B. Co-Citation of Sources

In this database, there are a total of 3595 cited sources. By keeping the threshold of a minimum of 15 citations per cited source, a total of 35 were able to accomplish the threshold.

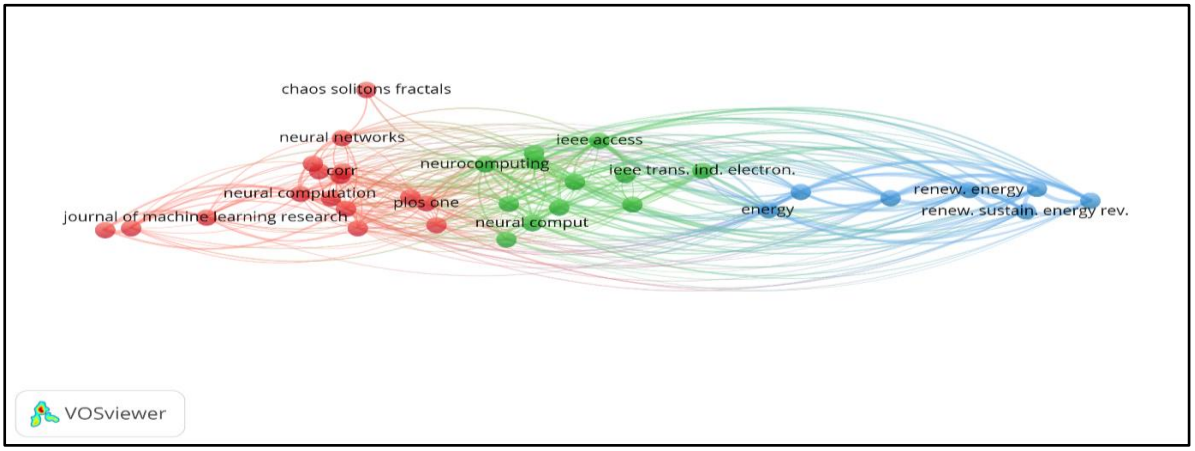

Figure 18: Network Analysis showing Co-Citation of Sources

Limitations of this study

In this bibliometric survey, only the Scopus information base is considered for distributions having the blend of watchwords. Various information bases are available such as Google Scholar, PubMed, and Web of Science, for exploring the examination distributions. These information bases could also have been utilized for such explorations. Utilizing such vast data sets would present a different set of conclusions based on the distribution of references. Most of the analysis was done using the words in the English language. Certain specific catchphrases are utilized by the authors. Various blends of the catchphrases, by adding or eliminating the words that have significance can be used further to explore the database for publications. Subsequently, this assessment overview has an expansion for future work to be done.

Conclusion

Through the survey carried out in this bibliometric study, it is evident that the use of LSTM networks for multivariate time series forecasting has attracted the attention of researchers very recently. It also shows an increasing trend in the number of publications. The number of research papers published in 2019 was more than double the entire number of papers published within the preceding four years. This bibliometric survey is completed on the information drawn from research publication documents from the Scopus database, within the characterized time of 2015- 2021. The first quarter of 2021 has already witnessed publications of more research studies than in any year between and including 2015 and 2018. LSTM networks have been explored for various applications in several domains where multivariate time series were involved. This analysis thus indicates the suitability of LSTM networks for multivariate time series forecasting.

References

Alhirmizy, S., & Qader, B. (2019). Multivariate Time Series Forecasting with LSTM for Madrid, Spain pollution. *2019 International Conference on Computing and Information Science and Technology and Their Applications (ICCISTA)*, 1–5. https://doi.org/10.1109/ICCISTA.2019.8830667

Amin, P., Cherkasova, L., Aitken, R., & Kache, V. (2019). Automating Energy Demand Modeling and Forecasting Using Smart Meter Data. *2019 IEEE International Congress on Internet of Things (ICIOT)*, 133–137. https://doi.org/10.1109/ICIOT.2019.00032

Chen, B., Lin, P., Lai, Y., Cheng, S., Chen, Z., & Wu, L. (2020). Very short-term power prediction for PV power plants using a simple and effective RCC-LSTM model based on short-term multivariate historical datasets. *Electronics (Switzerland)*, *9*(2). https://doi.org/10.3390/electronics9020289

Li, S., Xie, G., Ren, J., Guo, L., Yang, Y., & Xu, X. (2020). Urban PM2.5 Concentration Prediction via Attention-Based CNN–LSTM. *Applied Sciences*, *10*(6), 1–17. https://doi.org/10.3390/app10061953

Liu, F., Lu, Y., & Cai, M. (2020). A hybrid method with adaptive sub-series clustering and attention-based stacked residual LSTMs for multivariate time series forecasting. *IEEE Access*, *8*, 62423–62438. https://doi.org/10.1109/ACCESS.2020.2981506

Munkhdalai, L., Munkhdalai, T., Park, K. H. O., Amarbayasgalan, T., Batbaatar, E., Park, H. W. O. O., & Ryu, K. H. (2019). An end-to-end adaptive input selection with dynamic weights for forecasting multivariate time series. *IEEE Access*, *7*, 99099–99114. https://doi.org/10.1109/ACCESS.2019.2930069

Santra, A. S., & Lin, J. L. (2019). Integrating long short-term memory and genetic algorithm for short-term load forecasting. *Energies*, *12*(11), 1–11. https://doi.org/10.3390/en12112040

Tran, N., Nguyen, T., Nguyen, B. M., & Nguyen, G. (2018). A Multivariate Fuzzy Time Series Resource Forecast Model for Clouds using LSTM and Data Correlation Analysis. *Procedia Computer Science*, *126*, 636–645. https://doi.org/10.1016/j.procS.2018.07.298

Wang, K., Li, K., Zhou, L., Hu, Y., Cheng, Z., Liu, J., & Chen, C. (2019). Multiple convolutional neural networks for multivariate time series prediction. *Neurocomputing*, *360*, 107–119. https://doi.org/10.1016/j.neucom.2019.05.023

Wang, Y. S., Gao, J., Xu, Z. W., Luo, J. D., & Li, L. X. (2020). A prediction model for ultra-short-term output power of wind farms based on deep learning. *International Journal of Computers, Communications and Control*, *15*(4), 29–38. https://doi.org/10.15837/IJCCC.2020.4.3901

Wang, Y., Xie, D., Wang, X., & Zhang, Y. (2018). Prediction of wind turbine-grid interaction based on a principal component analysis-long short term memory model. *Energies*, *11*(11). https://doi.org/10.3390/en11113221

Wojtkiewicz, J., Hosseini, M., Gottumukkala, R., & Chambers, T. L. (2019). Hour-ahead solar irradiance forecasting using multivariate gated recurrent units. *Energies*, *12*(21), 1–13. https://doi.org/10.3390/en12214055

Wu, S. (2020). Research on Wind Power Ultra-short-term Forecasting Method Based on PCA-LSTM. *IOP Conference Series: Earth and Environmental Science*, *508*(1), 0–8. https://doi.org/10.1088/1755-1315/508/1/012068

Yu, Y., Cao, J., & Zhu, J. (2019). An LSTM Short-Term Solar Irradiance Forecasting under Complicated Weather Conditions. *IEEE Access*, *7*, 145651–145666. https://doi.org/10.1109/ACCESS.2019.2946057

Zhu, Q., Li, H., Wang, Z., Chen, J. F., & Wang, B. (2017). Short-term wind power forecasting based on LSTM. *Power System Technology*, *41*(12), 3797–3802.