KVS Ramachandra Murthy et al. (Eds.)

doi:10.3233/ATDE221245

Bibliometric Analysis on Identifying Plant, Crop Diseases Using Machine Learning and Deep Learning

Ch.V. Raghavendran[1], RVVN Bheema Rao[2], SK Mahaboob Basha[3], T.R. Mani Chigurupati[4]
[1,2]*Aditya College of Engineering and Technology, Surampalem, AP, India*
[3, 4] *TKR College of Engineering & Technology, Hyderabad, Hyderabad, Telangana, India*
[*1]raghuchv@yahoo.com, [2]rvvnbrao@hotmail.com, [3]bashashaik1617@gmail.com, [4]tulasiratnamani.chigurupati@gmail.com

Abstract. This paper is intended to explore the research done on identifying the diseased plants and crops using Machine Learning (ML) and Deep Learning (DL) techniques during last 10 years using bibliometric methods. In this study, we used Scopus database to analyze on "Plant disease" or "Crop disease" using "Machine Learning" or "Deep Learning" or "Neural Networks". This paper focuses on the importance of ML and DL techniques in identifying plant or crop diseases. The database collected from the Scopus is analyzed using VOSviewer software of version 1.6.16. The study is limited to publications from conferences, journals with subject areas are limited to Computer Science, Engineering and languages limited to English and Chinese. Scopus search outputs 824 articles on Plant or Crop diseases with ML, DL and Neural Networks covering conference papers and journal articles. Statistics showed that more articles were published during the last five years and major contributions were from India. By analyzing database on Authors, Subject area, Keywords, Affiliation, Source type it is evident that there is plenty of research scope in this area. Network analysis on diverse parameters specifies that there is a good scope to do research in this topic using advanced deep learning techniques.

Keywords. Crop Disease, Plant Disease, Machine Learning, Deep Learning, Neural Networks, Network Analysis, Bibliometric Analysis.

1. Introduction

Diseases in plant, crop are the major concern in the agricultural sector, and their automatic detection is critical for their monitoring. Plants are extremely susceptible to seasonal illnesses, which worsen over time and under varying climatic conditions. As a result, it is critical to address the issue of safeguarding plants from a variety of illnesses. The leaves reflect the majority of disease symptoms; however leaf identification by professionals in laboratories is costly and time-consuming. Rice infections commonly cause yield losses of 20 to 40% and are closely associated to the global economy [1]. Rice disease diagnosis is still primarily done by hand. It's vital to identify diseases quickly so that treatment can be planned and losses can be minimized.

Many computer algorithms have been developed to detect plant diseases early in order to protect crops from damage. Extracted features are very important in both segmentation and categorization of infected areas in the Machine Learning, Deep

Learning domains [2,3]. Deep learning approaches have recently expanded their applicability in plant disease identification, providing a comprehensive instrument with extremely accurate findings. Since Convolutional Neural Networks (CNNs) have achieved outstanding results, deep CNN models are used to categorize and analyze diseases in plants from the leaves [4-8].

In this context, this paper presents a thorough assessment of the literature with the goal of determining the current state of the art in the application of ML, DL and CNN in the identification and classification of plant diseases, as well as identifying trends using Scopus database and VOSViewer software.

2. Literature study

In previous decades, the citation count of a paper was regarded the most essential indicator of its scientific effect, and it was regularly used to evaluate the work of faculty members, research institutions, and universities [9]. Counts are not simply a measure of influence, but they can also be used to scale the overall quality of research [10].

A deep CNN based modified LeNet proposed in [11] to carry experiments on maize leaf images of PlantVillage dataset. This CNN is trained on four classes and achieved an accuracy of 97.89%. In the work [12], the researchers described a test that was done in Sardinia using the DSS LANDS to forecast potato late light illness. The objective of the study was to see if using a Machine Learning technique, regional weather characteristics could be utilized to forecast potato late blight risk in southern Sardinia. The disease severity is forecasted with 96% accuracy using feed forward neural network with parameters provided by ARPAS weather stations.

Citrus fruits quality and yield will decrease the “Citrus” disease. An effective segmentation and classification model for citrus plant disease is presented in the research [13]. Back Propagation Neural Network (BPNN) is used to categorize citrus disease using the selected features. The proposed model has been tested on the CDIG data and it outperformed other approaches and achieved superior outcomes on the test images.

Identifying earliest symptoms of infections through visual examination or laboratory studies is one of the known strategies for minimizing loss due to plant illnesses. Using an infrared camera and an artificial neural network for image classification, researchers presented a low-cost, portable, and noninvasive plant health monitoring device in [14]. The first prototype was tested on a cassava plant to determine its photosynthetic activity, which is a good indicator of the plant’s current health.

For the needed analysis, various analytical models such as Decision Tree, Random Forest, and Bayesian Neural Network are existing in the literature. These methods are applied for examining multidimensional, time-specific data in the agriculture segment in order to provide useful knowledge to help economic growth efforts [15-21].

3. Database

Data about the research work and publications can be obtained from popular repositories like Scopus, Web of Science, Scimago, Google Scholar etc. Among these Scopus database is the most popular and commonly used repository. Different keywords are used in the search and languages limited to English and Chinese. Another filter is on the year of publication with range May, 2021 to 2012.

4. Statistical analysis and results

4.1 Documents by year of publication, year by source

The documents collected for this analysis are from the years 2012 to May, 2021 from the Scopus database including Conferences, Book chapters, Articles etc. The following graph in Figure 1 shows progressive growth in the selected area.
The documents under analysis are collected from different sources viz., Springer Lecture Notes, ACM, IEEE etc. Among the list of sources, *Computers and Electronics In Agriculture* stands first with 44 documents and *Advances In Intelligent Systems And Computing* with 35 documents. The Figure.2 shows the graph representing year wise count with source of documents for the top five sources.

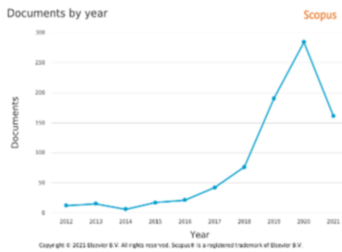

Figure. 1 Analysis of Documents by year

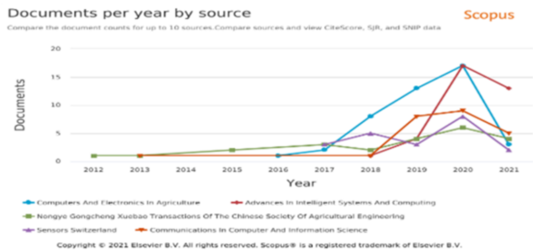

Figure. 2 Analysis of Documents by source

4.2 Documents by affiliation, Country / Territory

In analyzing the documents by the affiliations, it was found that *Chinese Academy of Science* has 22 documents and Vellore Institute of Technology with 14 documents. According to Scopus database Indian researches have published 325 documents, Chinese researches published 178 documents.

4.3 Documents by type, Subject

Among documents under study, it is observed that majority of the papers are Conference papers and Journal articles. Out of 824 resultant documents 746 documents are of these two types only. Doughnut chart presenting major documents types is shown in Figure. 3. The Scopus database has resulted 1987 documents for the given query and this study is limited two subjects Computer Science, Engineering. These two subjects consist of 1160 documents in which Computer Science is with 691, Engineering with 469. The pie chart in Figure. 4 presents the details of all the subjects with percentage of documents.

5. Network Analysis

5.1 Co-authorship analysis

In the Network analysis, the Co-authorship analysis is carried on three parameters – Authors, Organizations and Countries.

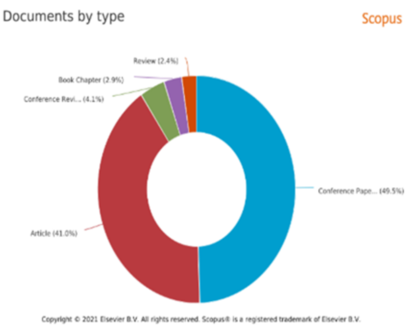

Figure. 3 Analysis of Documents by type

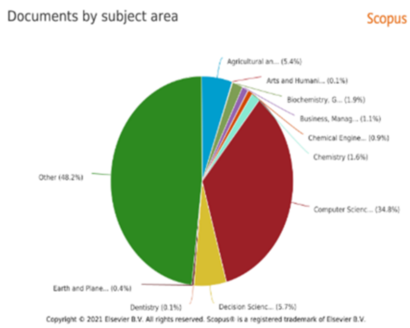

Figure. 4 Analysis of Documents by subject

In analyzing the Co-authorship with authors, documents with more than 25 authors are not considered. The min. no. of documents an author is made to two and these selected 356 authors from the total 2308 authors. Few of 356 items in the network are not connected to one another and the major set of connected items contains of 144 items only. It is witnessed from the network diagram that, Zhang J. has highest no. of documents equal to 13. The Figure. 5 shows Co-authorship relationship with authors.

In the selected documents, the researchers are from 1401 organizations and threshold set for network analysis is minimum number of documents per organization as two. This has resulted 71 documents and out of these only four items are connected in two clusters.

The researchers participated in the selected domain are from 87 different countries and we set a threshold of minimum documents of country two. Out of 48 documents, the largest set of connected items consists of 43 items with 15 clusters and 80 links.

5.2 Network Analysis of Co-occurrences

In this Network analysis, the Co-occurrences analysis is carried on three parameters – all keywords, Author keywords, Index keywords.

Keyword plays a vital role in identifying a document. Among the documents under study, there were 4490 keywords. For a threshold of minimum 3 occurrences, 775 documents are selected. The keyword *Deep Learning* stands first with 668 links with total link strength of 4143. The next keyword is *Plant disease* with 672 links and total link strength of 3440. The network with 775 documents is presented in Figure 6.

For the co-occurrence with author keywords the threshold is taken as minimum three occurrences of keywords and these resulted 176 documents out of 1600. In these *Deep Learning* stands first with 121 links and link strength of 572.

5.3 Network Analysis of Citations

Among 824 documents under study, 366 documents are selected with a threshold of two. In these, the largest set of connected items is 163 with 29 clusters and 307 links. It is observed that, the document titled *Deep Learning models for plant disease detection and diagnosis* of *Ferentinos K.P.* is having highest number of 64 links and 498 citations.

Figure. 5 Authors Network Analysis

Figure. 6 All Keywords Network Analysis

6. Conclusion

This study is on the role of Machine Learning, Deep Learning and Neural Networks in the field of identifying diseases in plants and Crop. This study is carried by using the popular Scopus database. Statistical analysis on the database is made by using Scopus and Network analysis is by using VOSViewer software. A total of 824 documents are selected by Scopus for the years from 2012 to 2021.

During this analysis different parameters are considered for comparison. From the year wise number of publications it is very clear that much research is going in this field. Considering documents from an organization, Chinese Academy of Sciences is on top with 22 documents. In the 824 selected documents, major contributions are from India viz., 325 followed by China with 178. Major no. of articles are published in the Conference proceedings and then in Journals and this contributed about 90% of total articles. Scopus database search has resulted 1987 documents for the given query and in this 691 documents are from Computer Science field contributing 34.8%.

Network analysis is done using VOSViewer software. Using this software, analysis is done on co-authorship in combination with Authors, Organizations and Countries. Then Co-occurrences are analyzed with All keywords, Author keywords and Index keywords. During Citation analysis; documents, sources, authors, country, and organization are considered. Bibliographic coupling analysis included documents and authors as unit of analysis. From this analysis it is concluded that major research is happening in the field of plant, crop disease using Machine learning and Deep learning using neural networks.

References

[1] Yibin Wang, Haifeng Wang, Zhaohua Peng, Rice diseases detection and classification using attention based neural network and bayesian optimization, Expert Systems with Applications, Volume 178, 2021, 114770, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2021.114770.

[2] Kianat, Jaweria & Khan, Muhammad & Sharif, Muhammad & Akram, Tallha & Rehman, Amjad & Saba, Tanzila. (2021). A Joint Framework of Feature Reduction and Robust Feature Selection for Cucumber Leaf Diseases Recognition. Optik. 240. 166566. 10.1016/j.ijleo.2021.166566.

[3] Nguyen, Ty & Paik, Incheon. (2020). Feature Extraction with Triplet Loss to Classify Disease on Leaf Data. 1-5. 10.1109/iCAST51195.2020.9319494.

[4] André Abade, Paulo Afonso Ferreira, Flavio de Barros Vidal, Plant diseases recognition on images using convolutional neural networks: A systematic review, Computers and Electronics in Agriculture, Volume 185, 2021, 106125, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2021.106125.

[5] Puangsuwan, Thararat & Surinta, Olarik. (2021). Enhancement of Plant Leaf Disease Classification Based on Snapshot Ensemble Convolutional Neural Network. ICIC Express Letters. 15. 669-680. 10.24507/icicel.15.06.669.

[6] Ghiani L, Sassu A, Palumbo F, Mercenaro L, Gambella F. In-Field Automatic Detection of Grape Bunches under a Totally Uncontrolled Environment. Sensors. 2021; 21(11):3908. https://doi.org/10.3390/s21113908

[7] Jadhav, S.B., Udupi, V.R. & Patil, S.B. Identification of plant diseases using convolutional neural networks. Int. j. inf. tecnol. (2020). https://doi.org/10.1007/s41870-020-00437-5

[8] K. Li et al., "Classification of different degrees of ginkgo leaf disease based on deep learning," 2019 6th International Conference on Systems and Informatics (ICSAI), 2019, pp. 1228-1232, doi: 10.1109/ICSAI48974.2019.9010110.

[9] Safer, M. A., & Tang, R. (2009). The psychology of referencing in psychology journal articles. Perspectives on Psychological Science, 4(1), 51–53.

[10] Anderson, M. H. (2006). How can we know what we think until we see what we said?: A citation and citation context analysis of Karl Weick's the social psychology of organizing. Organization Studies, 27(11), 1675–1692.

[11] Ahila Priyadharshini, R., Arivazhagan, S., Arun, M. et al. Maize leaf disease classification using deep convolutional neural networks. Neural Comput & Applic 31, 8887–8895 (2019). https://doi.org/10.1007/s00521-019-04228-3

[12] Fenu, Gianni & Malloci, Francesca. (2019). An Application of Machine Learning Technique in Forecasting Crop Disease. 76-82. 10.1145/3372454.3372474.

[13] C. Senthilkumar and M. Kamarasan, "Optimal Segmentation with Back-Propagation Neural Network (BPNN) Based Citrus Leaf Disease Diagnosis," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), 2019, pp. 78-82, doi: 10.1109/ICSSIT46314.2019.8987749.

[14] Panganiban, Edward. (2019). Cassava leaf NDVI - Artificial Neural Network (CaNDVI-ANN): A Low Cost, Portable and Non-Invasive Cassava Plant Health Monitoring Device. International Journal of Advanced Trends in Computer Science and Engineering. 8. 3555-3559. 10.30534/ijatcse/2019/136862019.

[15] S. V. Militante, B. D. Gerardo and N. V. Dionisio, "Plant Leaf Detection and Disease Recognition using Deep Learning," 2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), 2019, pp. 579-582, doi: 10.1109/ECICE47484.2019.8942686.

[16] Nitin N. Patil, Mohmmad Ali M. Saiyyad, Machine Learning Technique for Crop Recommendation in Agriculture Sector, 2019 International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-9 Issue-1, pp. 1359-1363.

[17] A. Akhtar, A. Khanum, S. A. Khan and A. Shaukat, "Automated Plant Disease Analysis (APDA): Performance Comparison of Machine Learning Techniques," 2013 11th International Conference on Frontiers of Information Technology, 2013, pp. 60-65, doi: 10.1109/FIT.2013.19.

[18] Yu-Hui F. Yeh, Wei-Chang Chung, Jui-Yu Liao, Chia-Lin Chung, Yan-Fu Kuo, Ta-Te Lin, A Comparison of Machine Learning Methods on Hyperspectral Plant Disease Assessments, IFAC Proceedings Volumes, Volume 46, Issue 4, 2013, Pages 361-365, ISSN 1474-6670, ISBN 9783902823304, https://doi.org/10.3182/20130327-3-JP-3017.00081.

[19] Wei, Yong & Chang, RuoKui & Wang, Yuanhong & Liu, Hua & Du, Yanhong & Xu, Jianfeng & Yang, Ling. (2011). A Study of Image Processing on Identifying Cucumber Disease. IFIP Advances in Information and Communication Technology. 370. 201-209. 10.1007/978-3-642-27275-2_22.

[20] Amara, Jihen & Bouaziz, Bassem & Algergawy, Alsayed. (2017). A Deep Learning-based Approach for Banana Leaf Diseases Classification. 79-88.

[21] J. P. Shah, H. B. Prajapati and V. K. Dabhi, "A survey on detection and classification of rice plant diseases," 2016 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC), 2016, pp. 1-8, doi: 10.1109/ICCTAC.2016.7567333.