IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 May 2024, accepted 7 July 2024, date of publication 10 July 2024, date of current version 18 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3425897

SURVEY

Deep Learning in Leaf Disease Detection (2014–2024): A Visualization-Based Bibliometric Analysis

JYOTISMITA CHAKI[1], (Senior Member, IEEE), AND DIBYAJYOTI GHOSH[2]
[1]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore 632014, India
[2]VIT Business School, Vellore Institute of Technology, Vellore 632014, India

Corresponding author: Dibyajyoti Ghosh (dibyajyoti.ghosh@vit.ac.in)

This work was supported by the Vellore Institute of Technology (VIT), Vellore, India.

ABSTRACT The agriculture industry is critical to delivering high-quality food and contributes significantly to the growth of economies and people which can be affected by the plant disease. This article demonstrates a visualization based bibliometric analysis to depict research trends in deep learning-based leaf disease detection from 2014 to January 2024. The publications used in this study are collected from the Scopus database. The research distributions with respect to sources and country, research trends, and research limits for deep learning in leaf disease detection studies are presented using Biblioshiny and VOSViewer software and visualization technologies. From 2014 to January 2024, the literature on this field has grown at an average rate of 53.41%. 1307 peer-reviewed publications from 54 countries are identified that are published in 594 distinct sources. India is the most productive country, accounting for 36.6% of total publications and 23% of total citations. Chitkara University Institute of Engineering and Technology was the most productive research institute, with 66 publications and 291 citations, while Computers and Electronics in Agriculture journal has the most citations in deep learning-based leaf disease detection research. The findings, in particular, show that ''Convolution Neural Network'', ''Transfer Learning'', ''Ensemble Learning'', etc., are the most widely used research topics in this field from 2014 to January 2024, and the research interest engrossed on applications of deep learning standard architectures. This study gives an insight into deep learning in leaf disease detection's general research patterns, which may assist researchers better understand and forecast the field's dynamic paths.

INDEX TERMS Bibliometric study, biblioshiny, deep learning, leaf disease detection, VOSViewer.

I. INTRODUCTION

Agriculture helped to domesticate the majority of today's food crops and animals many thousands of centuries before. Food insecurity, which is a significant source of plant diseases, is one of the most serious global issues confronting humanity today. Agriculture sector contributes to the largest percentage of economic growth in every country [1]. As a result, plant disease and pest infections in agriculture may have an influence on the global economy by lowering food production quality. Prophylactic medicines are ineffective at preventing epidemics and endemics. Early detection and diagnosis of plant disease, along with an effective crop protection system, can help to avert production quality losses.

Identifying different forms of plant disease is incredibly essential and considered a critical issue. Early detection of plant disease may allow for more informed decision-making in agricultural output. Infected plants typically exhibit visible stains or patches on their stems, fruits, leaves, or flowers. More specifically, every disease and pest situation produce distinct patterns that may be utilized to identify problems. Identifying a plant disease needs skill and personnel. Early and accurate disease detection and diagnosis may lessen the risk of further plant harm. The diagnosis and categorization of these disease have become significant issues. Farmers' traditional methods for forecasting and detecting plant leaf

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

diseases can be tedious and inaccurate. Manually predicting the sorts of disease may cause problems. The failure to detect and categorize plant disease immediately may lead to the demise of agricultural plants, resulting in a considerable drop in output. Also, this may result in the use of an inappropriate medicine during the evaluation of the plant disease, reducing crop quality and harming the environment. Farmers that utilize computerized image processing tools in their fields can cut losses while increasing output [2].

Since infected spots are originally recognized as spots and patterns on leaves, there are several strategies to tackle detection challenges for plants as computer vision has evolved [3]. Researchers have developed many methods for reliably detecting and classifying plant diseases. Advanced technology may identify leaf diseases at an early stage, reducing their detrimental impacts. Machine Learning (ML) and Deep Learning (DL) are being widely explored for autonomous detection and diagnosis of plant diseases, as human monitoring is time-consuming and laborious. The application of ML and DL to plant disease identification has grown in popularity, with promising findings in reliably diagnosing plant diseases from digital images. Traditional ML approaches, like feature extraction and categorization, are widely employed in the area of plant disease recognition [4]. These approaches use image attributes including shape, texture, and color to build a classification model that can distinguish between diseased and healthy plants. These approaches are extensively utilized to detect diseases such as leaf rust, leaf blotch etc. along with disease signs caused by abiotic pressures like nutrient deficiency and drought [5], but they have drawbacks in precisely detecting delicate disease symptoms and detecting early-stage diseases. Furthermore, they struggle to interpret complicated, high-resolution visuals.

Convolutional neural networks (CNNs) have recently been proposed as DL algorithms for detecting plant diseases [6]. These approaches entail a network training to understand the fundamental properties of the images, allowing for the detection of delicate disease indicators that standard image processing approaches may miss [7]. DL architectures can deal with complicated and large sized images; hence they are appropriate for images with a high resolution [8]. However, these approaches need a substantial quantity of labelled data for training and may not be appropriate for unknown conditions. Additionally, DL architectures are computationally costly, which may limit their use in certain applications.

Now-a-days, multiple researches have presented several DL techniques for detecting leaf diseases. Though, most research have concentrated on a single disease type or plant variety. More study is required to produce a robust and generalizable architecture that can be used to a variety of plant types and diseases. There is also a demand for additional publicly existing datasets to train and evaluate DL architectures. One of the current advancements in this area is transfer learning (TL), which permits pre-trained architectures to be reused on fresh datasets. TL [9] and ensemble approaches [10] have recently developed as prominent strategies in applying DL to identify leaf diseases. Transfer learning is the process of refining pre-trained architectures on a given dataset to improve the performance of DL architectures. Ensemble approaches, in contrast, combine numerous architectures to increase overall performance while decreasing reliance on a single architecture. These techniques were used to improve the resilience and accuracy of leaf disease detection models. Furthermore, it can avoid overfitting, a typical issue in DL architectures in which the architecture achieves good accuracy on training data but unable to produce satisfactory results on validation or test data. Another important factor to study is the utilization of image augmentation methods, which include artificially increasing the size of a dataset by applying random modifications to the images. This strategy has been utilized to promote data variety and lessen reliance on vast amounts of labelled data.

Efficient evaluation of large amounts of scientific articles in a study topic is crucial in today's information age. Numerous researches have been conducted on DL for leaf disease identification, with notable successes in classification and pattern recognition in three dimensions such as: single DL model (without TL) [11], TL based DL approach [9], [12] and ensemble DL model [10], [13]. In this work, we used both bibliometric study and visualization to analyse the research tendencies in DL based leaf disease detection with respect to abovesaid three dimensions. Bibliometric analysis is the use of quantitative statistics to analyze information in books, papers, and other publications [14]. Bibliometric study offers several benefits over typical literature evaluations and summaries the overall trend of the research in a field. First, bibliometric approaches analyze thousands of papers in a database to create a comprehensive picture of a study topic, such as DL based leaf disease detection, using professional software and high-performance computers. Second, citation analysis in bibliometric approaches allows for quantitative measurement of the effect of a study area, researchers, and publications. Third, it is possible to rapidly identify research hotspots and classical literature in a certain topic.

Numerous bibliometric analyses have been done to examine research trends across numerous themes and disciplines, including artificial intelligence in healthcare [15], DL in finance [16], DL in text classification [17], cybernetics [18] etc. Li et al., [19] conducted a bibliometric analysis on the evolution of DL during 2007 – 2019 showcasing the involvement of deep leap learning in different fields. To our knowledge, no proper bibliometric analyses have been reported on the research of DL based leaf disease detection.

We conducted a bibliometric review by using bibliometric visualization techniques to summarize and evaluate articles, identifying trends and hotspots in the area of DL based leaf disease detection. We collected articles from the Scopus database. The main motivation of using only the Scopus database for our bibliometric analysis is as follows. Scopus is selected due to its comprehensive coverage of relevant journals and conferences in the field of computer science and agriculture. Scopus is a popular choice for bibliometric reviews due to its wide coverage of scholarly publications across disciplines. It offers strong citation tracking and analysis tools, making it efficient for researchers to gather and analyze publication data. Bibliometrics is associated with the study of research using citation counts and patterns, which allows for the discovery of trends in research patterns as well as the quantitative evaluation of research outcomes. Consequently, many of these measures may be identified via the Web of Science. We used DL and leaf disease-related keywords to define search queries. Two primary research questions have been selected: (a) What are the strategic themes around DL based leaf disease detection? (b) What are the key challenges and possibilities for DL based leaf disease detection? The answers to these questions are studied and deliberated using bibliometric methods. This study offers a foundation for future research, including the following contributions:

1) Analyzation of the number of papers published in the area of DL based leaf disease detection from 2014 to 2024 (January).
2) Identification of the journals with the most publications, the most widely used techniques in DL based leaf disease detection, and the countries from where most of the authors are contributed in this domain.
3) Find the most commonly cited related articles in the literature.
4) Identification of the frequency of the author's keywords and the keyword cooccurrence analysis.
5) Identification of hot research topics in DL based leaf disease detection for future studies.
6) Identification of the predominant deep learning models used in leaf disease detection from 2014 to 2024.
7) Identification of countries and institutions have contributed most significantly to this field.

This study provides assistance for researchers interested in the potential of DL based leaf disease detection. It identifies the most active researchers and recent key research subjects among the authors. This study focused on DL rather than ML, as DL gained popularity for image detection purposes compared to ML. DL is expected to remain a valuable method for image detection, based on current advances.

The remaining article is divided into four sections. In Section II, we outline the methodology, including the article selection method and analytical tool utilized in this study. Section III presents results and discussion which includes statistical data on the distribution of DL based leaf disease detection studies by year, country, organization, author, source publication, research keyword etc. Section V compares the current study with an existing bibliometric study in the field of DL in agriculture. Section V discusses the challenges and future research direction in DL based leaf disease detection research. Section VI concludes the article and provides suggestions for further study in this topic.

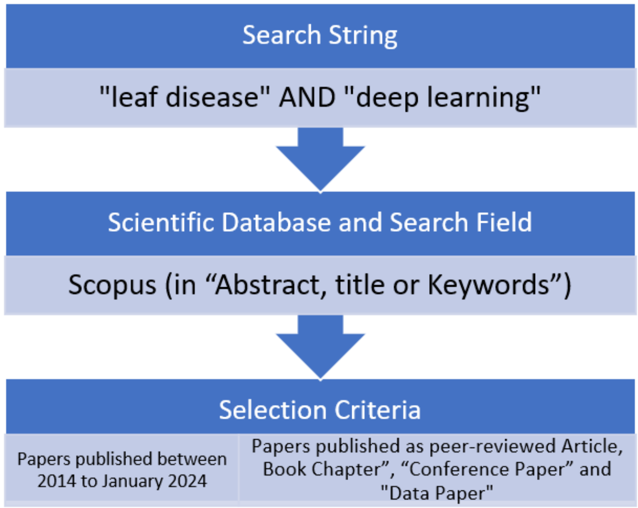

FIGURE 1. **The first phase of publication selection method.**

II. METHODOLOGY

A. ARTICLE SELECTION METHOD

The exploration for literature of this study was conducted on January 25, 2024, utilizing the Scopus database. The Scopus database was searched for articles containing the keywords "Deep Learning" and "Leaf Disease" in the title, abstract, or author's keywords, with a timeframe of "2014 to January 2024" and document type of "Article", "Book Chapter", "Conference Paper" and "Data Paper". The findings are selected in two formats: "plain text" and "comma-separated full record". The subsequent data was gathered for each article: title of the paper, abstract, source title, author, keywords, country, institution, funding details, publisher, and references. This search scope, search string, and selection criteria were defined as the first phase of the paper selection method as shown in Fig. 1

The search criteria in Fig. 1 returned 1597 publications. After the initial step, publications were divided into two categories based on the level of work required for analysis. Publications that just required reading the title and abstract were included in Group 1. The criteria used to exclude publications from further analysis in this category were as follows:

1) Publications not published in English.
2) Publications that are unrelated to the research objective, such as arguments on plant disease concern.

Publications in Group 2 required a careful study of their content. The criteria used to exclude publications from further analysis in this category were as follows:

1) Publications that did not include leaf disease and/or DL approaches.
2) Publications that focused on constructing a software tool or system rather than using DL approaches to leaf disease detection.

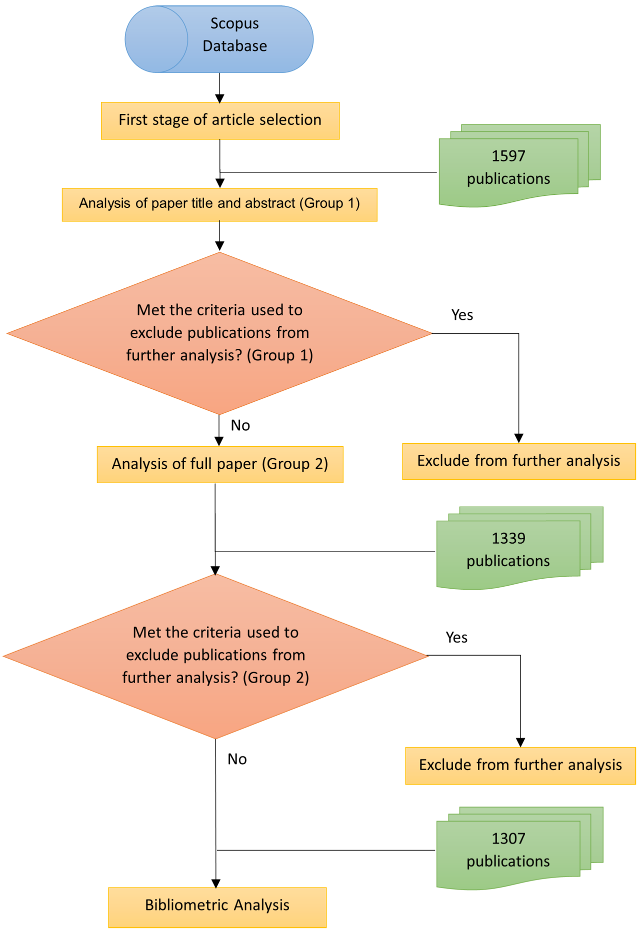

FIGURE 2. **The processes utilized for the bibliometric study of literature.**

3) Publications whose entire texts were unavailable for download.
4) Publications with overly particular applications, such as DL for plant disease detection or ML for leaf/plant disease detection.

The flowchart in Fig. 2 highlights the processes utilized for the bibliometric study of literature in this work.

Fig. 2 illustrates that 1307 publications are selected for further investigation and considered during the bibliometric study.

B. ANALYTICAL TOOL

Two main analysis tools were used to conduct the bibliometric analysis: biblioshiny and VOSViewer. The main motivations for using the two aforementioned analysis tools are as follows. Biblioshiny 4.0 from RStudio is an open-source bibliometric program. Biblioshiny's superiority stems from the fact that, unlike other bibliometric tools, it offers a comprehensive set of statistical techniques and visualizations that allow for performance analysis and conceptual mapping of the topic of research [55]. Furthermore, Biblioshiny is released as an open-source R package with a web-based graphical interface. The UI is straightforward and well-organized. The menu includes statistics and graphs for four levels of metrics (source, author, document, and Clustering by Coupling), as well as three types of knowledge structures (Conceptual Structure, Intellectual Structure, and Social Structure). The analysis possibilities are extensive and are classified into seven categories: 1) Introduction, 2) Sources, 3) Authors, 4) Documents, 5) Conceptual Structures, 6) Intellectual Structure, and 7) Social Structure. VOSviewer is a software application for creating and viewing bibliometric networks [56]. These networks may contain journals, researchers, or individual articles, and they may be built via citation, bibliographic coupling, co-citation, or co-authorship relationships. VOSviewer also has text mining capabilities, which allow the creation and display of co-occurrence networks of key phrases taken from scientific literature.

III. RESULTS AND DISCUSSION

A. OVERALL PUBLICATION TREND

The quantity of academic publications on a subject is a reliable predictor of publishing patterns, reflecting the research process in that discipline. After selecting 1307 publications, related to the DL based leaf disease detection, for bibliometric analysis, we were able to divide them by year of publication, resulting in the time series visualization displayed in Fig. 3. Fig. 3 indicates a steady increase in leaf disease detection using DL studies over the past few years, with an average yearly growth rate of 53.41%. The number of publications grew from 43 in 2019 to 592 in 2023. Till January 2024 the number of publications in this domain is 40. From 2020 to January 2024, 1256 papers have been published, accounting for 96.1% of all DL-based leaf disease detection publications from 2014- January 2024. This suggests that DL-based leaf disease detection research has developed substantially from 2020. A recent surge in publications (2020 onwards) could be fueled by multiple forces. One driver could be the remarkable progress in deep learning. These algorithms can analyze massive datasets with unprecedented speed and accuracy, potentially leading to a wave of new research and publications across various disciplines. Additionally, specific global priorities might be influencing publication trends. For example, a growing population and anxieties about food security could be driving a rise in publications focused on improving agricultural productivity. Essentially, the recent surge in publications might be a confluence of advancements in research tools like deep learning and a global focus on pressing issues like food security.

Fig. 3 displays the annual publishing growth rate and acceleration. There were no publications related to DL-based leaf disease detection from 2014 – 2016. Over the past 4.1 years (2020 – January 2024), the average annual growth velocity has been 306 articles per year, with a peak acceleration in 2020. In 2022 and 2023, the published literature on this domain grew by 374 and 592, respectively, indicating a surge in research activity. This is mostly due to the diffusion of DL methodologies and electronic databases.

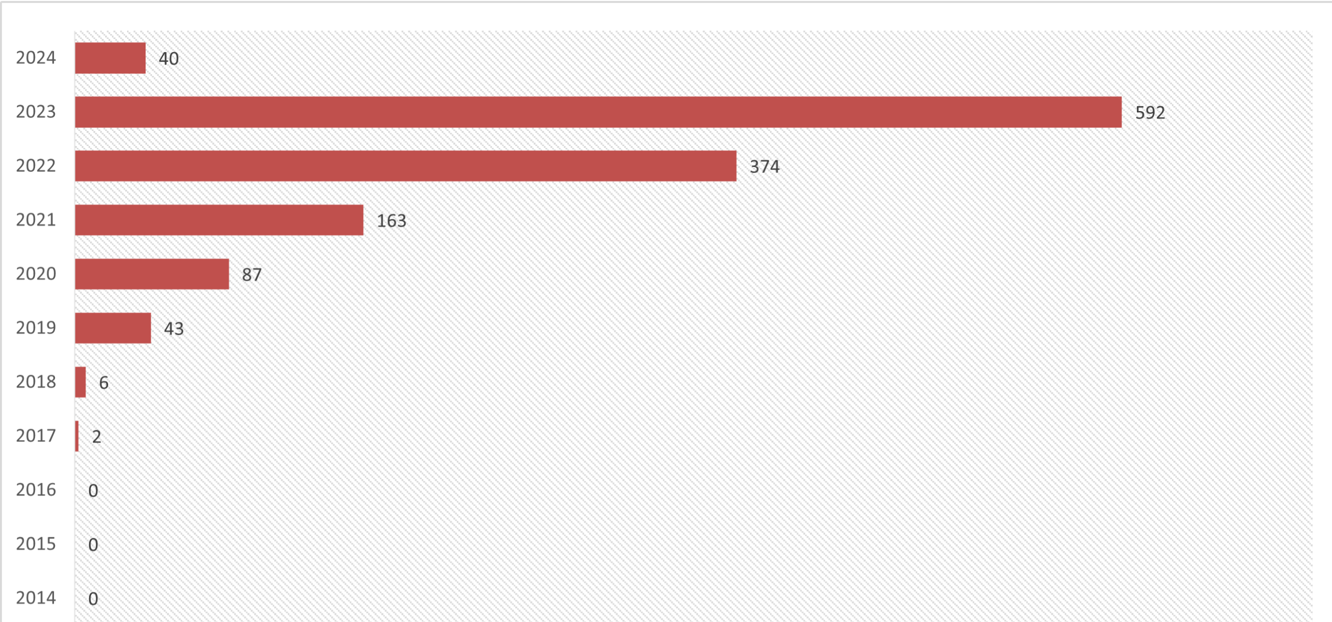

FIGURE 3. DL-based leaf disease detection publication rate from 2014 – January 2024.

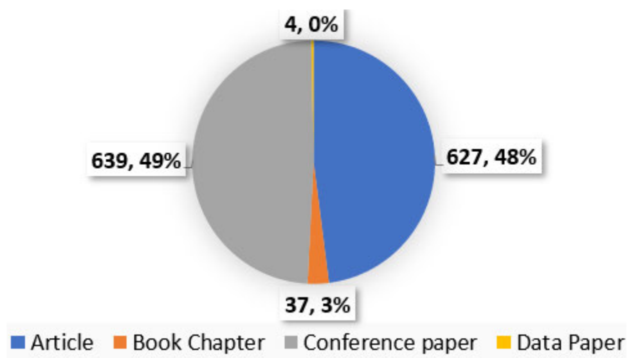

FIGURE 4. The source-wise publications in DL-based leaf disease detection.

TABLE 1. Ten most productive countries in DL-based leaf disease detection.

Country	Total no. of publications	Percentage	Total Citations	Percentage
India	479	36.6	3600	23
China	371	28.4	5150	32.8
Bangladesh	168	12.9	155	1
Saudi Arabia	43	3.3	286	1.8
Korea	33	2.5	151	1
Turkey	21	1.6	534	3.4
Egypt	20	1.5	300	1.9
Pakistan	19	1.4	168	1.1
Malaysia	15	1.1	462	2.9
Indonesia	13	0.9	54	0.3

Fig. 4 denotes the source-wise publications in DL-based leaf disease detection. From the figure, it can be noted that the maximum papers are published in either journal (627) or conference proceedings (639).

B. COUNTRIES AND RESEARCH INSTITUTES

Biblioshiny's research found 1307 publications on DL-based leaf disease detection studies from 54 different countries. Table 1 includes the ten most productive countries, which account for 90.4% of all publications (1307 publications). India has the most papers published on DL-based leaf disease detection, with 479 publications and 3600 citations, accounting for 36.6% and 23% of the whole, correspondingly. China ranks second with 371 (28.4%) research papers and 5150 citations (32.8%), followed by Bangladesh with 168 (12.9%) publications and 155 (1%) citations. This follows seven countries that were ranked from 4th to 10th, with 12.5% of publications and 12.5% of citations on average. One interesting point to be noted is even if Saudi Arabia, Turkey, Egypt, Pakistan, and Malaysia have a smaller number of publications compared to Bangladesh these countries have a greater number of citations than Bangladesh. One reason for having more citations with a smaller number of publications can be the collaboration or outreach of the publications done by Saudi Arabia, Turkey, Egypt, Pakistan, and Malaysia. This is also applicable to India and China. Overall, we can conclude that India and China have dominated DL-based leaf disease detection studies.

To examine country co-authorship for DL-based leaf disease detection publications, Fig. 5 shows a network map built using VOSviewer for collaboration across 22 countries with at least 10 publications. The map displays a country's impact in the research area (total number of publications) based on node size, while the thickness of linkages indicates cooperative proximity between countries. Table 2 shows detailed information related to the 10 countries that collaborated most in DL-based leaf disease detection research. The

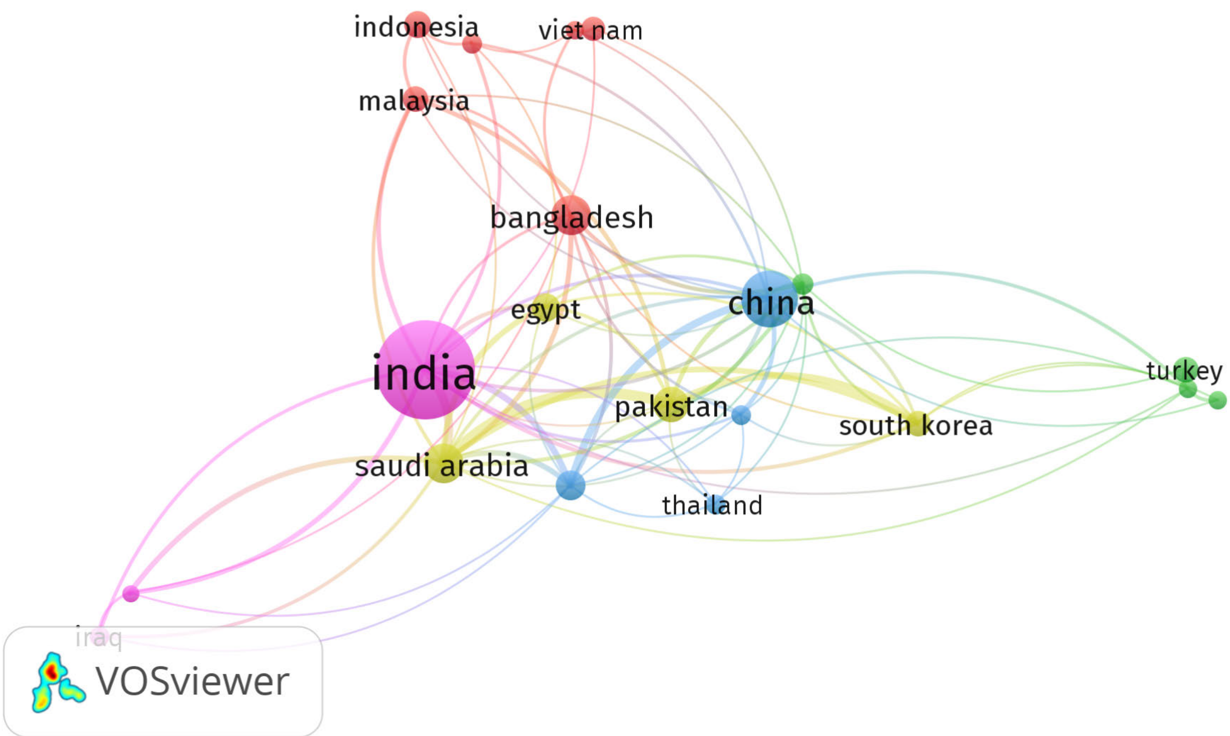

FIGURE 5. A network map visualization demonstrating the collaboration across 22 countries with at least 10 DL-based leaf disease detection publications.

TABLE 2. The detailed information related to the 10 countries that collaborated most in DL-based leaf disease detection research.

Country	Total number of collaborations	Total No. of Documents
India	16	725
Saudi Arabia	15	74
China	14	185
United Kingdom	13	17
United States	11	39
Pakistan	10	54
South Korea	10	26
Bangladesh	9	74
Egypt	8	33
Australia	8	15

top four countries i.e., India, Saudi Arabia, China, and the United Kingdom have the highest number of collaborations. Even if Saudi Arabia published a smaller number of papers compared to China, it has more collaborations (compared to China) with other countries in DL-based leaf disease detection research. Another interesting observation from Table 2 is even if Saudi Arabia and Bangladesh have published the same number of papers in DL-based leaf disease detection, the collaboration of Bangladesh is not as strong as Saudi Arabia.

We also examined the most significant research institutions for DL-based leaf disease detection studies. There were 1094 organizations engaged in the retrieved papers, with Table 3 highlighting the top 10 institutions. The 10 institutions' publishing numbers are more or less evenly distributed. DL-based leaf disease detection papers published by Northeast Agricultural University from China have more citations (905) than other organizations, despite only publishing 16 papers in this area. This suggests that publications from this institute are more significant. This statistical data indicates that India has a substantial effect on DL-based leaf disease detection research, with 9 out of the top 10 organizations based there. On the contrary several institutions, like Zayed University, Yangzhou University, etc. are decreasing their focus on DL-based leaf disease detection

TABLE 3. The top 10 most significant research institutions for dl-based leaf disease detection studies.

Organization Name (Country)	No. of publications	No. of Citations
Chitkara University Institute of Engineering and Technology (India)	66	291
SRM Institute of Science and Technology (India)	31	155
Kongu Engineering College (India)	27	307
Daffodil International University (Bangladesh)	26	191
Vellore Institute of Technology (India)	26	485
Lovely Professional University (India)	21	297
Koneru Lakshmaiah Education Foundation (India)	19	64
Veltech University (India)	17	44
Graphic Era Hill University (India)	16	35
Northeast Agricultural University (China)	16	905

research and have published only one paper from 2014 to January 2024.

Fig. 6 depicts collaboration amongst DL-based leaf disease detection research institutions. Each node represents an institution, with the connections representing collaboration. The size of the nodes and width of the links is dependent on the connection with the institute's publication count and collaboration closeness, correspondingly. In Fig. 6, the minimum number of papers for an organization is set at 10. Thus, 23 institutions met the criterion. Chitkara University Institute of Engineering and Technology has the most collaboration with other research organizations, with 66 links and a link strength of 3766, after that Lovely Professional University (links = 21, link strength = 2247) and University of Engineering and Technology Taxila (links = 14, link strength = 2119), indicating their active involvement in DL based leaf disease detection research.

C. ANALYSIS OF CO-AUTHORSHIP

Analysis of co-authorship can aid scientists in identifying collaboration possibilities and providing insights into research networks. We found that 3957 authors contributed to 1307 publications in DL-based leaf disease detection, with an average of 3.99 authors per publication. In this regard, 3290 authors (83.1% of the total) produced just one publication, while 428 authors (11%) wrote two, and 118 authors (3%) published three papers. VOSViewer analyzes co-authorship by constructing a network visualization of researchers (Fig. 7). In Fig. 7, the minimum number of citations and publications for a researcher is set at 5. Finally, 33 researchers met the criterion. Each node signifies a researcher, with larger nodes indicating more publications. Links represent the authors' cooperative closeness. The analysis found that individual researchers make up the bulk of productive authors in DL-based leaf disease detection research, with low co-authorship.

D. IMPORTANT SOURCE

The 1307 retrieved documents were published across 594 sources, with an average of 2.2 publications per journal. Of the 594 sources, Lecture Notes in Networks and Systems and Multimedia Tools and Applications journal (3.7%) published 48 papers, Frontiers in Plant Science journal (2.4%) published 32, and IEEE Access journal (2.2%) published 29 articles. Fig. 8 illustrates the trend of the source's production over time in the top five journals in the field of DL-based leaf disease detection research. Lecture Notes in Networks and Systems, Frontiers in Plant Science journal, IEEE Access journal, and Multimedia Tools and Applications journal have steadily increased since 2021. Computers and Electronics in Agriculture journal displayed a steady growth from 2019.

Table 4 shows the most productive journals for DL-based leaf disease detection research, organized by citation counts. The Computers and Electronics in Agriculture had the most citations (2310), despite only publishing 28 papers on DL-based leaf disease detection research. IEEE Access ranks second with 1730 citations, followed by Frontiers in Plant Science (412 citations), Multimedia Tools and Applications (201 citations), and Computers, Materials and Continua (160 citations). Both Computers and Electronics in Agriculture and Communications in Computer and Information Science Journal started their publication in DL-based leaf disease detection in 2019. But from their number of citations, we can conclude that Computers and Electronics in Agriculture journal have more outreach for the DL-based leaf disease detection publications compared to the Communications in Computer and Information Science journal even if the number of publications in these two journals is almost the same in this field. DL-based leaf disease detection research articles are published in multidisciplinary science publications, including computer science, medicine, civil engineering, mathematics, electrical engineering, and other areas.

TABLE 4. The most productive journals for dl-based leaf disease detection research, organized by citation counts.

Source Name (Country)	No. of publications	No. of citations	Start publication year
Computers and Electronics in Agriculture (Netherlands)	28	2310	2019
IEEE Access (United States)	29	1730	2018
Frontiers in Plant Science (Switzerland)	32	412	2020
Multimedia Tools and Applications (Netherlands)	48	201	2020
Computers, Materials, and Continua (United States of America)	21	160	2021
Communications in Computer and Information Science (Germany)	25	121	2019
Agronomy (Switzerland)	17	117	2021
Lecture Notes in Networks and Systems (Switzerland)	48	91	2021
Lecture Notes in Electrical Engineering (Germany)	25	38	2020
International Journal of Intelligent Systems and Applications in Engineering (Turkey)	14	18	2023

E. MOST RELEVANT KEYWORDS

The authors' keyword terms characterize the article's limitations and influence on the scientific community. Keyword analysis is the collection of keywords from relevant works in an area. The tool finds keywords, analyzes patterns, and highlights research directions. Table 5 shows the top 10 keywords. To remove discrepancies, we combined the semantically equivalent or comparable terms "CNN" and "convolutional neural network". From the table, we can conclude that Convolutional Neural Network, Transfer Learning, and Image Processing techniques are widely used by researchers in DL-based leaf disease detection research. The importance of these keywords in DL-based leaf disease detection research is discussed in detail in the Keyword

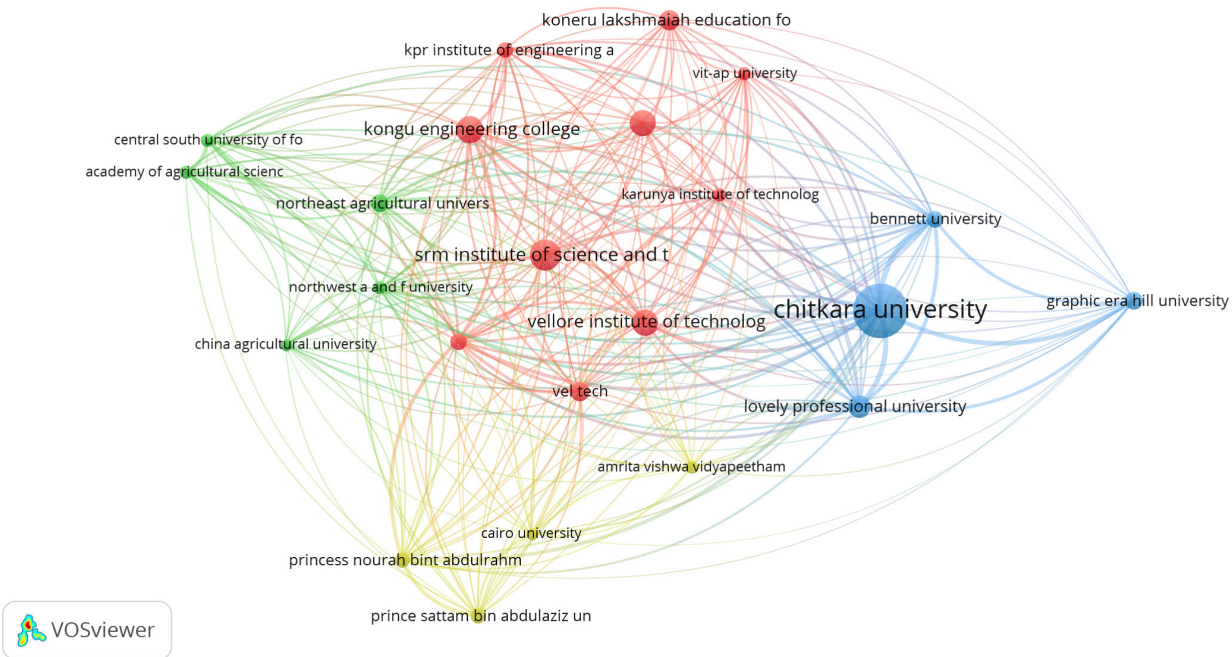

FIGURE 6. Collaboration amongst DL-based leaf disease detection research institutions.

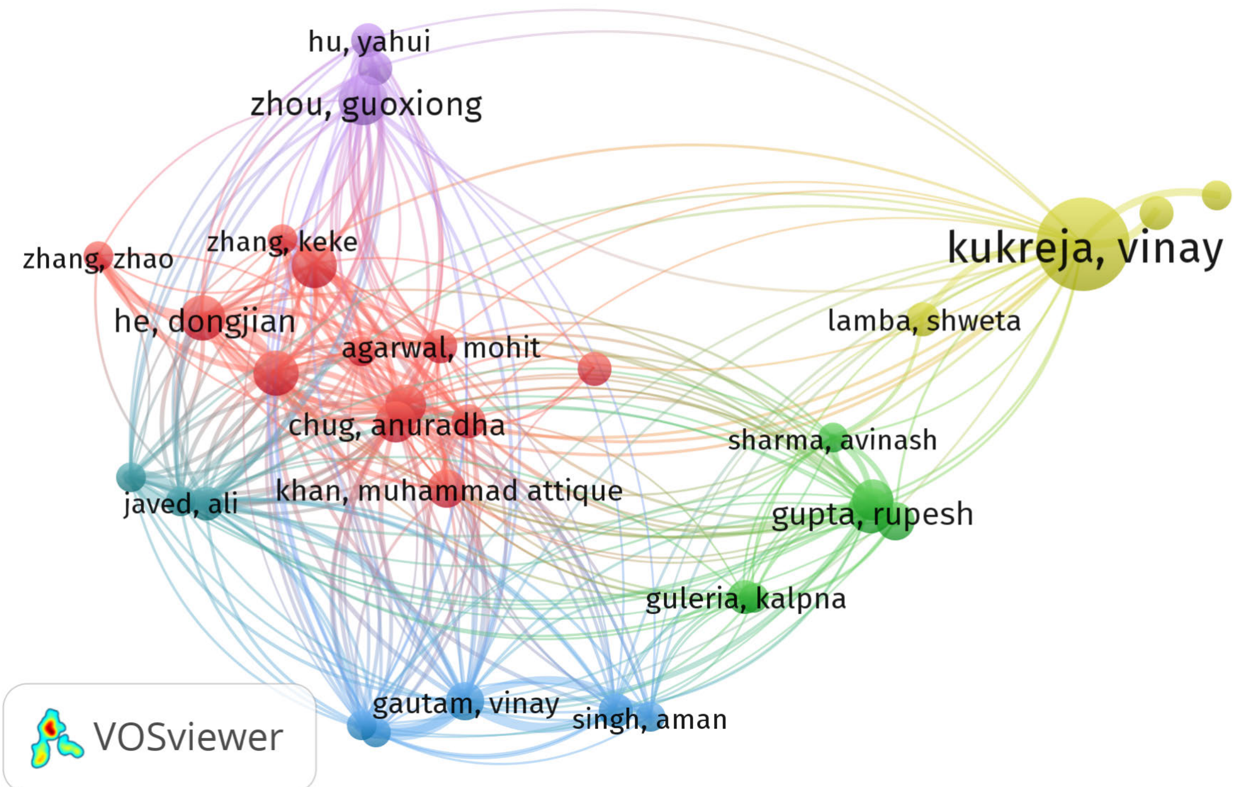

FIGURE 7. Co-authorship visualization.

co-occurrence analysis section. ML is another important keyword in this research field as many researchers have compared their DL-based technique with the ML-based techniques to demonstrate the superiority of DL-based methods in leaf disease detection research. For example, in [4], authors compare the results obtained from different

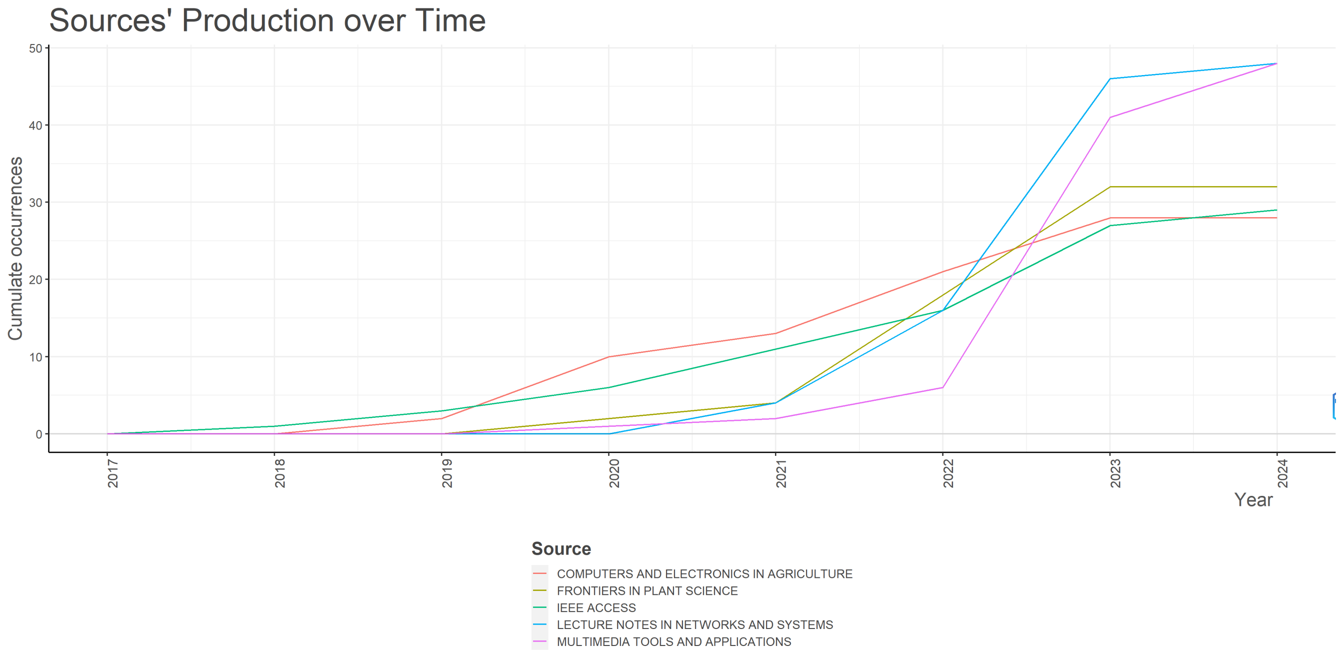

FIGURE 8. **The trend of the source's production over time in the top five journals in the field of DL-based leaf disease detection research.**

TABLE 5. **The top 10 keywords in dl-based leaf disease detection research.**

Keyword	Occurrences
Convolution Neural Network	556
Transfer Learning	189
Image Processing	113
Machine Learning	109
VGG16	53
Artificial Intelligence	50
Semantic Segmentation	50
Data Augmentation	44
Deep Neural Network	43
InceptionV3	41

MLs like Stochastic Gradient Descent (SGD), Random Forest (RF), Support Vector Machine (SVM), and DL (VGG19, VGG16, InceptionV3) for detecting citrus plant diseases. The disease categorization accuracy (CA) obtained through investigation and DL approaches outperform ML methods in the case of disease detection Technology reported CA are as follows: SGD (86.5%), SVM (87%), RF (76.8%), InceptionV3 (89%), VGG19 (87.4%), and VGG16 (89.5%). According to the results, RF provides the minimum CA, whereas VGG16 provides the top CA.

In this work, a word cloud visualization was used to collect data on the most popular author keywords, as seen in Fig. 9. The size of a chart element depends on the number of documents containing the keyword. These keywords refer to significant agricultural research areas using DL-based solutions. From the figure, we can conclude that transfer learning is one of the most recommended methodologies which includes VGGNet, AlexNet, ResNet, DenseNet, MobileNet, GoogleNet, InceptionV3, and XCeptionNet were mostly used by the researchers for the DL-based leaf disease detection.

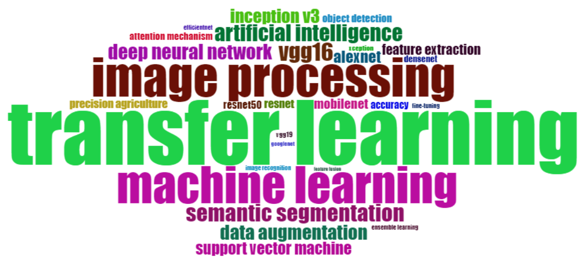

FIGURE 9. **A word cloud visualization of the most popular author keywords in DL-based leaf disease detection.**

F. KEYWORD CO-OCCURRENCE ANALYSIS

Co-occurrence refers to terms that are similar and related to the same issue, but not identical. Bibliometric analysis uses author keywords to identify research hotspots within a discipline. We use VOSviewer to create a keyword co-occurrence network for DL-based leaf disease detection, as shown in Fig. 10.

To begin, we combined keywords having similar meanings, such as ''deep neural networks'' and ''deep neural network'' or ''convolutional neural networks'' and ''convolution neural network''. The minimum number of keyword occurrences is set to 50. Out of 2111 saved keywords, 14 encountered the criterion. Fig. 10 shows that the 14 keywords were divided into 5 clusters (red, green, blue, yellow, and purple) to identify research hotspots using VOSViewer's default clustering algorithm. Table 6 displays the keyword and their occurrence rates within each cluster.

Keywords inside the same cluster share a comparable hotspot. Cluster 1 contains terms related to DL-based leaf disease detection studies, whereas clusters 2, 3, and 4 focus on benchmark techniques such as Convolution Neural Network, Image Processing, Ensemble Learning, Feature Fusion, and

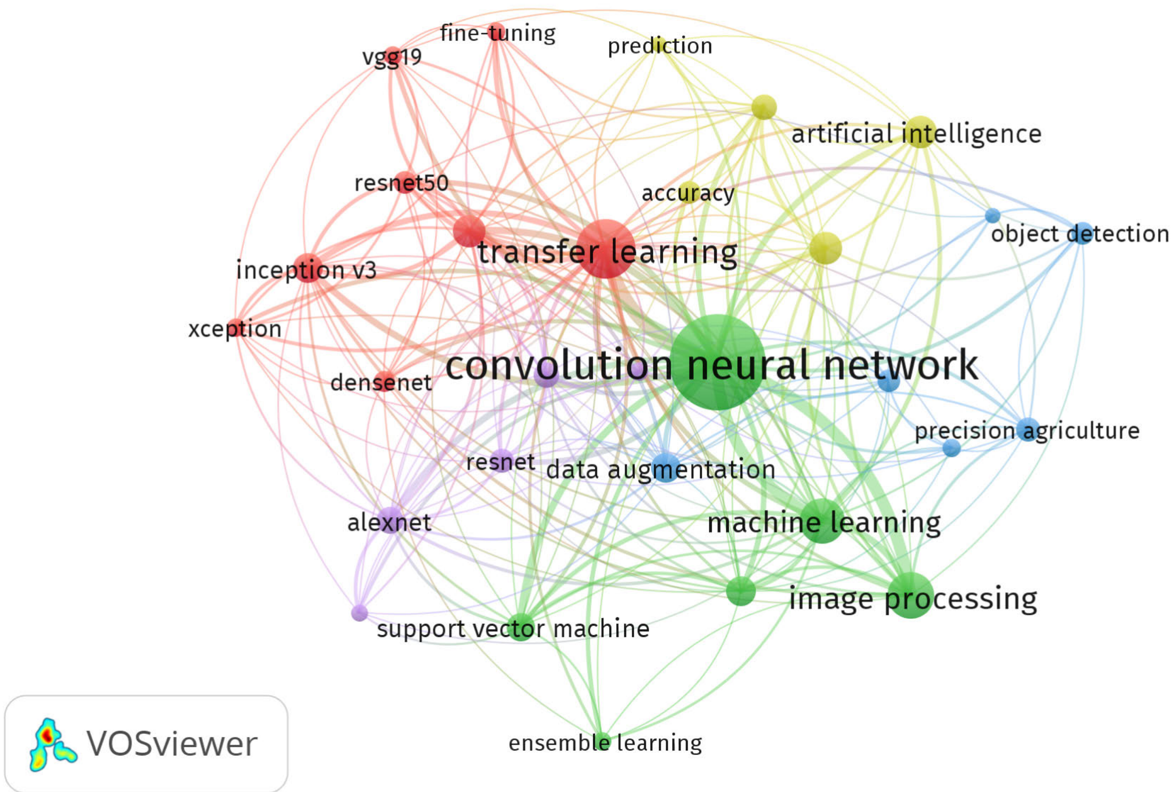

FIGURE 10. A keyword co-occurrence network for DL-based leaf disease detection.

TABLE 6. The keyword in dl-based leaf disease detection research and their occurrence rates within each cluster.

Cluster 1 (Red)	Cluster 2 (Green)	Cluster 3 (Blue)	Cluster 4 (Yellow)	Cluster 5 (Purple)
Transfer learning (189)	Convolution Neural Network (556)	Data Augmentation (44)	Artificial Intelligence (50)	AlexNet (35)
VGG16 (53)	Image Processing (113)	Precision Agriculture (26)	Semantic Segmentation (50)	MobileNet (30)
InceptionV3 (41)	Machine Learning (109)	Object Detection (24)	Feature Extraction (31)	ResNet101 (26)
ResNet50 (24)	Deep Neural Network (43)	Attention Mechanism (22)	Accuracy (23)	EfficientNet (14)
DenseNet (20)	Support Vector Machine (37)	Image Recognition (14)	Prediction (10)	GoogleNet (13)
VGG19 (16)	Ensemble Learning (14)	Feature Fusion (11)		
Xception (16)				
Finetuning (14)				

Semantic Segmentation, etc. Cluster 5 keywords mostly refer to different DL-based leaf disease detection techniques. The five groups are categorized based on the research level of DL-based leaf disease detection.

1) CLUSTER 1 (RED COLOR): KEYWORDS RELATED TO DL-BASED LEAF DISEASE DETECTION STUDIES

Cluster 1 keywords correspond with the topic term "DL-based leaf disease detection" reflecting the research domain of DL-based leaf disease detection investigations. The most common terms in this cluster are transfer learning, VGG16, and InceptionV3, with co-occurrence frequencies of 189, 53, and 41, respectively. DL employs a strong mechanism known as transfer learning. Transfer learning has enabled deep neural network training with less data by using the capacity to reuse previous models as well as their understanding of new situations. Two factors contribute to the success of DL-based leaf disease detection studies: (1) The transfer learning method's ability to solve nonlinear, multivariate, nonparametric classification or regression problems effectively. For example, in [20], authors present a transfer DL-based technique for detecting tomato leaf disease using the Conditional Generative Adversarial Network to produce synthetic images of tomato plant leaves. A DenseNet121 model is then trained on simulated and real images using transfer learning to classify tomato leaf images into ten disease groups. The model used in [21] has been thoroughly trained and tested using the PlantVillage dataset. The suggested technique classified tomato leaf images into five, seven, and ten groups with 99.51%, 98.65%, and 97.11% accuracy, respectively. (2) Recent years have seen a significant increase in data in DL-based leaf disease detection research domains. The literature on DL-based leaf disease detection investigations has several instances. For example, in [21], authors used transfer learning-based VGGNet for the classification of grape and tomato leaf disease images collected from the PlantVillage dataset with an accuracy of 98.4% and 95.7% respectively. The authors of [44] focused on the investigation and study of leaf disease detection using the TL – TL-InceptionV3 architecture and refinement of the architecture. A vast variety of the architecture accuracy tests are performed by training the architecture with various parameters. When the network Batch size and the learning rate are set to 100 and 0.01 respectively, the network's training and testing accuracy are maximized. The training and testing precision rate for crop disease image detection in the PlantVillage Dataset is 95.8% and 93% respectively.

2) CLUSTER 2 (GREEN COLOR): RESEARCH OF CONVOLUTIONAL NEURAL NETWORK, IMAGE PROCESSING, AND ENSEMBLE LEARNING IN DL-BASED LEAF DISEASE DETECTION

The top two keywords in this cluster are convolutional neural network (CNN) and image processing with occurrences of 556 and 113. CNN is the most common method for detecting leaf diseases. The study in [22] describes two distinct CNN architectures for recognizing the kind of disease in tomato leaves. The first architecture uses residual learning to discover important characteristics for categorization. The second design uses an attention mechanism on top of the residual deep network. Experiments were done using the Plant Village Dataset, which included three diseases: leaf mold, late blight, and early blight. The suggested study used the attention mechanism to leverage features learned by the CNN at multiple processing hierarchies, resulting in an overall accuracy of 98% on the validation sets in the 5-fold cross-validation. A total of 15 cases have been fed to the CNN model in [23], out of which 12 cases are of diseased plant leaves namely, Potato Early Blight, Bell Paper Bacterial Spot, Tomato Target Spot, Potato Late Blight, Tomato Yellow Leaf Curl Virus, Tomato Mosaic Virus, Tomato Early Blight, Tomato Bacterial Spot, Tomato Leaf Mold, Tomato Late Blight, Tomato Spider Mites, and Tomato Septoria Leaf Spot 3 cases of healthy leaves, Potato Healthy, Bell Paper Healthy, and Tomato Healthy. The test accuracy is reached at 88.80%.

Digital image processing refers to the methods for modifying digital images using computer algorithms. Image processing is used to improve an existing image or to extract useful information from it. This is relevant in various DL-based Computer Vision applications, where such preprocessing can significantly improve model performance [24]. The study reported in [25] investigates the impact of various preprocessing methods like Adaptive Histogram Equalization, contrast limited adaptive histogram equalization (CLAHE), Image Sharpening, class imbalance methodologies like Synthetic Minority Oversampling Technique, Major-to-Minor Translation, Generative Adversarial Networks and DL classifiers in the plant disease detection pipeline. The evaluation results show that CLAHE combined with image sharpening and a generative adversarial network-based approach for resampling outperformed other preprocessing and resampling techniques, with an average classification accuracy of 97.69% and an average F1-score of 97.62% when fed through a ResNet50 DL classifier. In [26] researchers have used different image pre-processing methods like image cropping, resizing, adjusting brightness, adjusting contrast, flipping, denoising, and rotation before feeding the tomato leaf disease images to the ResNet50 classifier for the leaf disease detection purpose.

Other than these two, ensemble learning also has played an important role in DL-based leaf disease detection. The use of ensemble approaches can increase the accuracy of image detection models, regardless of the underlying methodology. Ensemble approaches integrate predictions from numerous models to get a final result [27]. For example, in [28] authors deployed three DL models commonly used in leaf disease identification approaches, namely VGG11, ResNet18, and MobileNetV3, and built an ensemble-based model employing voting and average weighting algorithms. 4 categories of leaves are used in this study and the accuracy achieved was reported as 98.1%, 95.8%, 98.4%, and 90.75% respectively by using the ensemble method which is greater than standalone DL models. In [29], the authors offer a deep ensemble learning model for autonomous plant disease detection. The pre-trained models are enhanced by transfer learning. Overfitting can be combated using various augmentation techniques such as image enhancement, rotation, and scaling. This study provides a comprehensive taxonomy of the performance of a single model and many ensemble learning models in classifying super-resolution tomato plant leaf disease photos. The planned prototype's efficiency is assessed using a publicly available dataset that includes 10 distinct biotic disease classes of tomato plants. The efficiency of a single pre-trained model, VGG16, is 98%. In addition, an ensemble of three models (VGG16, InceptionV3, and GoogleNet) produces a more accurate accuracy of 97.3%.

3) CLUSTER 3 (BLUE COLOR): RESEARCH OF DATA AUGMENTATION AND PRECISION AGRICULTURE IN DL-BASED LEAF DISEASE DETECTION

Besides image preprocessing and CNN, another major step of DL-based leaf disease detection is data augmentation with an occurrence of 44. Deep artificial neural networks in computer vision require a huge amount of training data to learn properly, which is costly and time-consuming to obtain. Data augmentation solves this problem by artificially increasing the training set via label-preserving modifications [30]. In [31] a novel network model, Fast WDBlock-based GAN, was developed to accomplish unsupervised data augmentation of tomato leaf disease images. In [32], augmented plant leaf disease datasets were created using basic image manipulation and DL-based image augmentation techniques such as image cropping, flipping, color transformation, rotation, noise injection, PCA color augmentation, Generative Adversarial Networks (GANs), and Neural Style Transfer (NST). The performance of the data augmentation strategies was evaluated using cutting-edge transfer learning algorithms such as VGG16, ResNet, and InceptionV3. A thorough simulation demonstrates that the augmented dataset utilizing

GAN and NST approaches outperforms the original dataset using a simple image manipulation-based augmented dataset.

Precision agriculture technology and techniques are another vital technology used in DL-based leaf disease detection with an occurrence of 26. This technology assists farmers in reducing the quantity of inputs they use, potentially lowering the risk of environmental damage. Precision agriculture technology can help farmers monitor and manage their crops more successfully, lowering the risk of pests and diseases and improving soil quality [33]. The purpose [34] is to identify and classify grape and mango leaf diseases using an 8,438-image dataset of damaged and healthy leaves sourced from the Plant Village dataset and locally obtained. The deep CNN is trained to detect disease or its absence. A pre-trained CNN architecture known as AlexNet is modeled for automated feature extraction and classification, achieving detection accuracy rates of 99% and 89% for grape leaves and mango leaves, correspondingly. Reference [35] proposes a classification technique that combines DenseNet121 with support vector machines to identify plant leaf diseases. The Plant-Village dataset was utilized to experiment. The new model's performance was evaluated using sugarcane plant leaves, and it achieved 97.78% classification accuracy above the previous DenseNet121-based classifier model (94%).

4) CLUSTER 4 (YELLOW COLOR): RESEARCH OF SEMANTIC SEGMENTATION IN DL-BASED LEAF DISEASE DETECTION

Semantic segmentation is a DL method that assigns a label or category to each pixel in an image. It is used to identify a set of pixels that represent separate categories [36]. For example, the diseased part on the leaf. Thus, semantic segmentation is another essential step in DL-based leaf disease detection with the occurrence of 50. In [37] authors proposed two CNN models for the classification of leaf disease: F-CNN considering full image and S-CNN considering the segmented disease portion of the leaf. The authors considered the tomato plant type with 10 classes taken from the Plant Village database. The average accuracy increases from 96.3% in the case of F-CNN to 98.0% in the case of S-CNN. In [38], the UNet and PSPNet networks were used to semantically segment leaf disease images from the BRACOT dataset, and a ResNet was utilized to classify the disease. The PSPNet and UNet networks have a mean intersection over the union of 93.5% and 94.3%, correspondingly.

5) CLUSTER 5 (PURPLE COLOR): RESEARCH ON DIFFERENT DL-BASED LEAF DISEASE DETECTION TECHNIQUES

Other than VGG16 and InceptionV3, the other most popular DL-based models used for DL-based leaf disease detection are AlexNet, MobileNet, and ResNet101 with the occurrence of 35, 30, and 26 respectively. In [39], the authors deploy the AlexNet architecture on the Android platform to forecast tomato leaf disease based on images. The prediction model was created using a dataset including 18,345 training and 4,585 testing data. The data is organized into ten labels for tomato leaf diseases. The best model using the Adam optimizer, 75 epochs, 128 batch size, and an uncompromising cross-entropy loss function has a top accuracy with a loss of 0.1331, a recall value of 0.99, an average of 98%, and an F1-count of 0.98, resulting in good and precise classification results. In [40], authors assessed and employed MobileNet architecture on one publicly available bean leaf disease dataset, which included two diseased classes (bean rust and angular leaf spot) and one healthy class. 1296 bean leaf images were used for the test purpose. The collected findings demonstrated that the MobileNet model obtains good categorization efficiency, with the average accuracy being more than 97% and 92% on the training and testing dataset. In [41] the author's dataset includes images of corn leaves with various disease severity. TL retrains a pre-existing ResNet architecture using the corn leaf dataset. The experimental findings show that the ResNet18, ResNet50, and ResNet101 architectures attained accuracy rates of 96.68%, 95.73%, and 95.26%, correspondingly. The ResNet101 architecture outperforms in terms of accuracy and recall measures.

G. KEYWORD, COUNTRY AND SOURCE RELATIONSHIP

India, China, and Saudi Arabia have active research communities on DL-based leaf disease detection. In Fig. 11, we use a Sankey diagram to visualize the primary components of three domains (e.g., countries, source, and keywords) and their relationships. The size of an element in a Sankey diagram represents its importance, which is proportional to its number of nodes. For example, the node size of the source depicts the number of DL-based leaf disease detection publications by the source. The study identifies the top sources (journals) where authors have published and discussed research subjects based on keywords in relevant countries.

From Fig. 11, it can be noted that China has published more articles in Frontiers in Plant Science, Computers in Electronics in Agriculture and Agronomy journal whereas Multimedia Tools and Applications, Lecture Notes in Networks and Systems, Communications in Computer and Information Science, Lecture Notes in Electrical Engineering and International Journal of Intelligent Systems and Applications in Engineering journals have more publications in DL based leaf disease detection from India.

H. TOP MANUSCRIPT BY CITATIONS

Document citation analysis assesses the impact of papers on the literature. In doing so, we conducted a citation analysis to identify the most significant works on the use of DL-based leaf disease detection and summarized their content. Table 7 shows the top ten most influential publications in terms of total citations (TC) and average citations per year (ACY).

Liu et al., [42] address the challenge of limited training data by creating additional diseased leaf images. This data augmentation helps the model learn more effectively. A novel CNN architecture based on AlexNet is proposed. This design achieves better performance with fewer parameters compared to the standard AlexNet, making it more efficient. The model, trained on a dataset of over 13,000 images, achieved an impressive 97.62% accuracy in identifying four common apple leaf diseases. Additionally, the generated images improved the model's accuracy by over 10%. These findings suggest that the proposed deep learning model offers a powerful solution for apple leaf disease detection. It boasts high accuracy, faster learning, and improved robustness thanks to the image generation technique.

Chen et al., [43] explores using transfer learning with deep CNNs for plant leaf disease identification. Instead of training a CNN from scratch, the authors leverage pre-trained models like VGGNet (trained on ImageNet) to jumpstart the learning process. These pre-trained models act as a strong foundation, reducing training time and potentially improving performance. The proposed method achieves impressive results. On a public dataset, it reaches a validation accuracy of over 91.83%. Even for images with complex backgrounds, the average accuracy remains high at 92.00% for rice leaf disease classification. This research suggests that transfer learning offers a powerful approach for accurate plant disease identification.

Jiang et al., [45] tackles real-time apple leaf disease detection using a novel deep learning approach. They build a robust dataset (ALDD) that incorporates both controlled lab images and complex field images. This dataset is further enhanced through data augmentation techniques to improve model generalizability. A new model named INAR-SSD is proposed. This model leverages the power of GoogLeNet's Inception structure and a unique ''Rainbow concatenation'' approach. This design allows the model to efficiently extract features from the images for accurate disease detection. The INAR-SSD model, trained on over 26,000 diseased leaf images, achieved a detection performance of 78.8% mAP (mean Average Precision) on the ALDD dataset. Notably, it also achieved a high processing speed of 23.13 frames per second (FPS). These results suggest that the proposed approach offers a promising solution for real-time disease detection in apple orchards, enabling faster and more efficient disease identification.

Zhang et al., [46] proposes two improved deep learning models for identifying maize leaf diseases with high accuracy and fewer parameters. The base architectures of GoogLeNet and Cifar10 were modified through techniques like parameter adjustment, pooling layer tweaks, adding dropout layers, and rectified linear unit (ReLU) functions. Additionally, they reduced the number of classifiers in the final layers. Compared to traditional models like VGG and AlexNet, these improved models boast a significantly smaller number of parameters. This leads to faster training times and potentially lower computational resources needed for deployment. The results were impressive. For eight maize leaf diseases, the improved GoogLeNet model achieved a top-1 accuracy (most likely class prediction) of 98.9%, while the Cifar10 model achieved an average accuracy of 98.8%.

Geetharamani and Pandian, [47] proposes a new deep learning model for identifying plant leaf diseases. The model utilizes a deep convolutional neural network (CNN) trained on a diverse dataset containing 39 classes of plant leaves and background images. To improve the model's performance, the researchers implemented several data augmentation techniques. By introducing these variations, the model learns to identify diseases from images with different orientations, lighting conditions, and minor imperfections. Furthermore, the training process involved experimenting with various training epochs, batch sizes, and dropout rates. This optimization helped the model achieve better performance compared to popular transfer learning approaches on the validation dataset. Ultimately, after extensive simulations, the proposed model achieved a remarkable classification accuracy of 96.46%. This suggests that the deep CNN with data augmentation is a promising approach for accurate plant leaf disease identification.

Atila et al., [54] investigated the effectiveness of EfficientNet, a deep learning architecture, for classifying plant leaf diseases. The researchers compared its performance to other cutting-edge models using the PlantVillage dataset. EfficientNet B5 and B4 achieved the highest accuracy (99.91% and 99.97% respectively) and precision (98.42% and 99.39% respectively) on both the original and augmented datasets, outperforming other models.

I. FUNDING SPONSOR

The survey selected the most productive and famous institutions that have supported the advancement of DL-based leaf disease detection research. Fig. 12 depicts the top 19 funding institutions and their related number of publications. China is the top to provide funds for 143 DL-based leaf disease detection publications. India comes next, by providing funds for 67 publications in this field. Saudi Arabia and Pakistan provided funds to publish 54 and 36 articles in DL-based leaf disease detection research.

IV. COMPARISON OF THE FINDINGS WITH EXISTING BIBLIOMETRIC REVIEW ARTICLE

A comparison of the findings from the current study is done with an existing bibliometric review article [57] to understand if the current observations align with the existing reviews. In [57], the bibliometric study is based on the evolution of DL in agriculture. In [57], authors show a jump from 74 publications in 2018 to 495 in 2021. The current study covers a broader timeframe (2019-2024), mentioning a rise from 43 publications in 2019 to 592 in 2023 with data for January 2024. Both studies depict a surge in research on DL-based leaf disease detection in recent years.

In [57], authors highlights ''Computers and Electronics in Agriculture'' (142 publications) as the most prolific journal, followed by ''Remote Sensing'' (95 publications) on DL based precision agriculture. It also mentions ''IEEE Access'' (34 publications) among the most cited sources. The current study identifies ''Lecture Notes in Networks and Systems'' and ''Multimedia Tools and Applications'' (with 48 papers each) as the top publishers, followed by ''Frontiers in Plant Science'' (32 publications) and ''IEEE

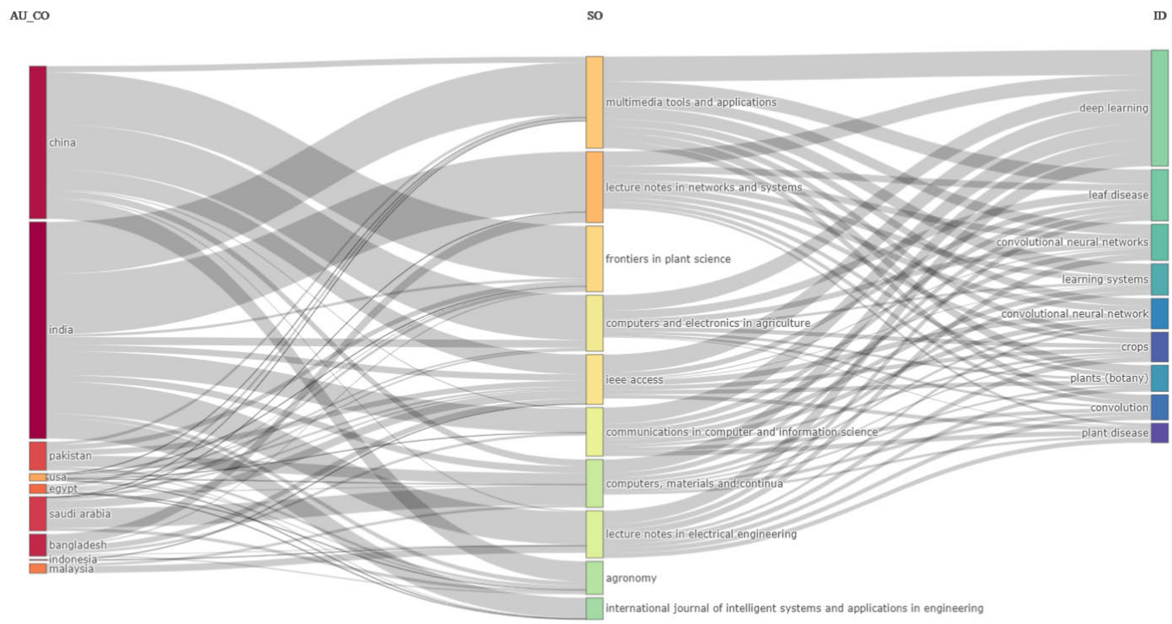

FIGURE 11. Sankey diagram to visualize the primary components of three domains (e.g., countries, source, and keywords) and their relationships in DL-based leaf disease detection research.

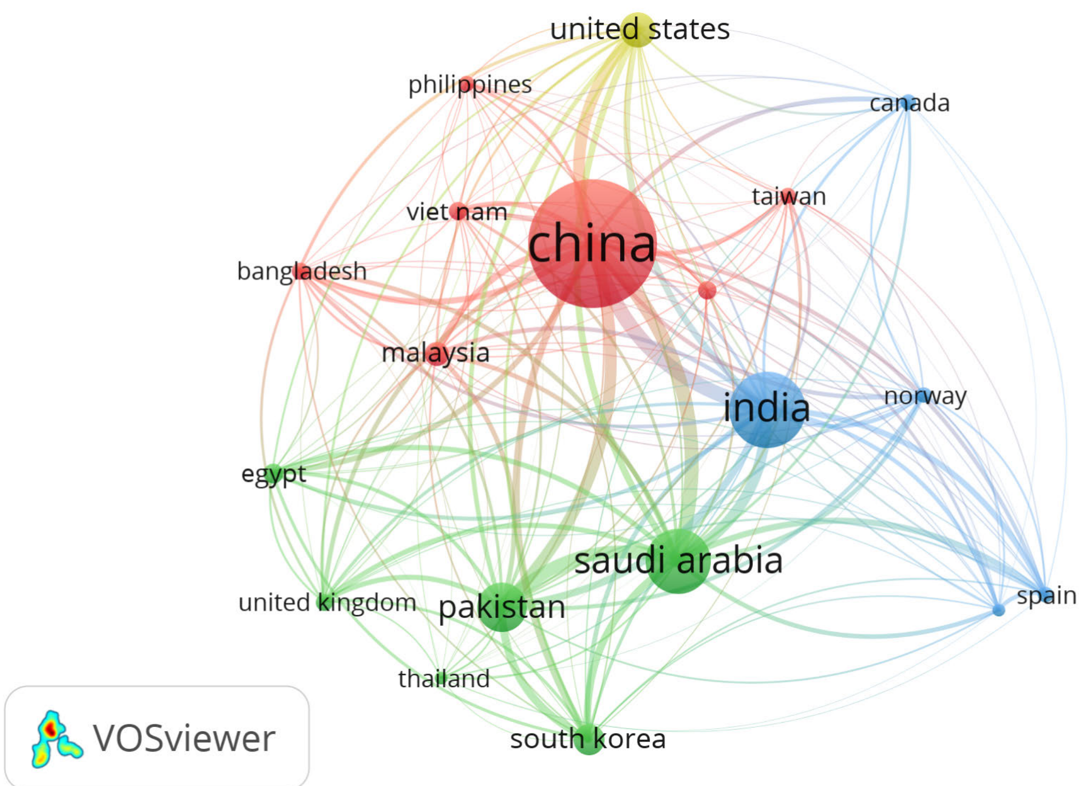

FIGURE 12. Top 19 funding institutions and their related number of publications in DL-based leaf disease detection research.

Access'' (29 publications). Both reviews reveal valuable insights into publication trends in agricultural research. In [57], authors highlight established journals across various precision agriculture subfields, while the current study delves deeper into specific journals publishing research on the fast-growing area of DL-based leaf disease detection.

The keywords identified in [57] showcases the paramount importance of deep learning, particularly CNNs, for image-based plant disease detection in agriculture. While other techniques like transfer learning, feature extraction, and segmentation might be important details, the core focus revolves around using deep learning models to classify and

TABLE 7. The top ten most influential publications in terms of total citations (TC) and average citations per year (ACY) in DL-based leaf disease detection research.

Paper	DOI	TC	ACY
LIU B, 2018, SYMMETRY	10.3390/sym10010011	500	71.43
CHEN J, 2020, COMPUT ELECTRON AGRIC	10.1016/ j.compag.2020.105393	473	94.60
JIANG P, 2019, IEEE ACCESS	10.1109/ ACCESS.2019.2914929	430	71.67
ZHANG X, 2018, IEEE ACCESS	10.1109/ ACCESS.2018.2844405	409	58.43
GEETHARAMANI G, 2019, COMPUT ELECTR ENG-a	10.1016/ j.compeleceng.2019.04.011	397	66.17
ATILA Ü, 2021, ECOL INFORMATICS	10.1016/ j.ecoinf.2020.101182	354	88.50
AMARA J, 2017, LECT NOTES INFORMATICS (LNI), PROC - SERIES GES INFORM (GI)	—	290	36.25
SUJATHA R, 2021, MICROPROCESSORS MICROSYST	10.1016/ j.micpro.2020.103615	248	62.00
SETHY PK, 2020, COMPUT ELECTRON AGRIC	10.1016/ j.compag.2020.105527	239	47.80
AGARWAL M, 2020, PROCEDIA COMPUT SCI	10.1016/ j.procs.2020.03.225	233	46.60

potentially localize plant diseases in images for practical applications in precision agriculture. The keywords identified in this study also highlights the paramount importance of deep learning, particularly CNNs, for plant disease detection. Image processing techniques are likely crucial for preparing the images for analysis by the deep learning models. Transfer learning might be a valuable approach to improve model performance. While machine learning algorithms in general and object detection are mentioned, the focus seems to be on deep learning-based methods for image classification and potentially object localization of plant diseases. So, we can conclude that in precision agriculture the most important part is to handle the plant disease which is the key interest of this study.

V. CHALLENGES AND FUTURE RESEARCH DIRECTION IN DL-BASED LEAF DISEASE DETECTION RESEARCH

Our bibliometric investigation shows that the DL method can accurately detect leaf disease images, comparable to manual detection. However, some factors must be addressed for consistent and trustworthy results. As DL gains popularity, there is growing interest in its use in agriculture.

A. LARGE IMAGE SIZE

In DL, image detection typically uses small-sized images as network inputs. Large images must be scaled to match the network requirements, as larger images demand more parameter estimates, memory, and processing power. Whole images are notoriously difficult to analyze in analysis, but shrinking the images may diminish cell information, resulting in less effective image detection. As a result, the whole image is frequently divided into patches (small sections) so that each one may be analyzed individually. Recently, Islam et al., [48] proposed a segmentation-based technique for detecting rice diseases from leaf pictures that makes use of deep neural networks. Disease-affected rice leaf areas were segmented using the local segmentation approach, and the CNN was trained using just this patch of images. The proposed technique has been applied to three distinct datasets, including one produced by us and consisting of rice leaf images acquired from the Bangladesh Rice Research Institute. Three cutting-edge CNN architectures, VGG, ResNet, and DenseNet, employed in the proposed technique, were trained on these three datasets to identify diseases. The suggested method's classification performance on the three datasets was evaluated and compared using the three CNN architectures mentioned above. These findings indicate that this model is extremely promising in categorizing rice leaf diseases. This approach may produce more accurate and reproducible imaging detection than human interpretation.

B. VARIATION IN IMAGE COLOR

Color variation in DL models can impact analytical outcomes. Variations in staining results can be attributed to factors such as staining solution of batch or manufacturer, leaf thickness, staining parameters, and scanner model [49]. DL models might suffer from poor performance if color variation is not included. Color augmentation, normalization, and grayscale conversion are some of the strategies proposed to address image color fluctuation [50]. Grayscale conversion is a straightforward procedure [51], but it may miss important information on color presentation. Color normalization involves changing an image's color values pixel by pixel utilizing techniques including color constancy, deconvolution, and transfer. Color normalization may be useful for images with identical cell or tissue compositions. Color normalization should be used with caution as it might affect the accuracy of leaf disease image detection algorithms [52].

C. DATA INADEQUACY

When there is inadequate data, CNN models tend to be less generic, which can lead to overfitting. One way to circumvent the problem is to use data augmentation tasks, which serve to improve the performance of CNN models in image detection. Recently many automated data augmentation techniques have been proposed, including multi-degree-of-freedom image capture. It is vital to evaluate the physical validity of the samples, as well as the impact of the various generated difficulties on algorithm performance. In recent years, many strategies have been used for producing synthetic data using generative adversarial networks. The generative adversarial network generates samples quickly in data augmentation applications, particularly image-to-image translation. Common issues in smart agriculture include the difficulty of getting high-quality disease samples and the high expense. To address this issue, [53] proposes a high-quality image augmentation approach based on a dual generative adversarial network (GAN). First, the original data were utilized to train Enhanced Wasserstein GANs, which produced augmented-data samples. The augmented data were fed into the Optimized-Real-ESRGAN, which produced high-quality augmented data samples. Lastly, the high-quality augmented-data samples were fed into the CNN for disease classification, and the method's efficacy was assessed using indicators.

D. APPLICATION OF UNSUPERVISED LEARNING

Leaf disease detection currently leans on supervised learning, demanding extensive labeled data. Future studies should explore the application of unsupervised learning techniques in leaf disease detection, an area that remains underexplored according to our findings. This method analyzes unlabeled leaf images, potentially uncovering hidden patterns of unknown diseases. Imagine a system that detects subtle changes in leaf color or texture, even for unseen diseases. Unsupervised algorithms can group similar leaf images, potentially revealing early signs of new outbreaks. This could revolutionize detection, empowering farmers and researchers to be proactive. Integrating unsupervised learning into disease detection frameworks can unlock a new level of identification and monitoring. This paves the way for adaptable agricultural practices, ultimately safeguarding crop yields and food security.

E. APPLICATION OF SEMI-SUPERVISED LEARNING

Leaf disease detection often struggles with limited labeled data for training models. One of the solutions can be the semi-supervised learning. This technique leverages both labeled and abundant unlabeled leaf images to enhance detection accuracy. Imagine a system that can learn from a small set of labeled diseased leaves, then utilize a vast pool of unlabeled images to refine its understanding of healthy and diseased states. Semi-supervised learning can exploit the inherent patterns within unlabeled data, improving disease classification. By incorporating semi-supervised learning, we can bridge the gap between limited labeled data and robust disease detection models. This approach offers a more efficient and scalable solution, ultimately benefiting farmers and researchers with improved crop health monitoring and yield protection.

F. LACK OF INTERDISCIPLINARY RESEARCH BETWEEN COMPUTER SCIENCE AND PLANT PATHOLOGY

Despite the potential of computer vision and deep learning for plant disease detection, a significant gap exists between computer science and plant pathology research. This lack of interdisciplinarity can be attributed to several factors. Firstly, domain expertise is often siloed. Computer scientists might struggle to understand the nuances of plant diseases, while plant pathologists may lack familiarity with deep learning techniques. Secondly, communication barriers can hinder collaboration. Different terminology and research objectives can create a disconnect between the two fields. The implications of this gap are significant. Missed opportunities exist for developing powerful disease detection tools. Additionally, slower progress can occur in both fields, as insights from each domain remain untapped. Bridging this gap through joint research efforts and educational programs would lead to significant advancements in automated disease detection and ultimately benefit agricultural productivity and food security.

VI. CONCLUSION

This study presents a visualization-based bibliometric analysis of DL-based leaf disease detection research trends across engineering areas, including methodologies and applications. The present study analyzed 1307 publications related to DL-based leaf disease detection published between 2014 and January 2024. We provide scientific maps depicting yearly publication numbers, countries, research institutes, author productivity, and source journals for DL-based leaf disease detection investigations. Analysis of document co-citation and keyword co-occurrence highlights research frontiers and hotspots in DL-based leaf disease detection. Also, the top manuscript based on citations and funding institutions is included in the study to understand the trend of this research and support from the institutions. At last, the challenges and future research directions are discussed to large image size, variation in image color, and data inadequacy. The key findings are as follows.

This study offers resources for researchers working on DL-based leaf disease detection studies, including selecting acceptable journals, identifying current research trends, and encouraging collaboration among authors and researchers. The collected high-frequency keywords assist researchers in identifying hotspots and understanding the research dynamics and directions.

DL-based leaf disease detection is a thriving research field, as evidenced by the increasing number of publications since 2020. The research distribution analysis shows that India and China have played the most major roles in DL-based leaf disease detection studies. India is the most productive country. Another interesting point that can be concluded from the study to country-wise number of publications and citations in DL-based leaf disease detection. Egypt, Turkey, and Pakistan published almost the same number of articles but compared to Pakistan and Egypt, Turkey has more influence in DL-based leaf disease detection research.

Although DL-based leaf disease detection publications are evenly distributed across research institutes, Chitkara University Institute of Engineering and Technology published the maximum number of articles in this domain than others. The Computers and Electronics in Agriculture journal has the highest influence among academic publications that publish DL-based leaf disease detection research.

DL-based leaf disease detection research's knowledge bases are retrieved through keyword analysis. Classical research papers include topics such as Convolution Neural

Networks, Transfer Learning, Image Processing, Machine Learning, Artificial Intelligence, Semantic Segmentation, VGG16, Data Augmentation, Deep Neural Networks, and InceptionV3. Our keyword co-occurrence analysis and clustering revealed that DL-based leaf disease detection research is mostly focused on DL benchmark methods and their applications. Furthermore, benchmark methods employed in DL-based leaf disease detection research often overlap with knowledge bases.

Journals with a focus on agriculture and computer science have a great influence on DL-based leaf disease detection research such as Computers and Electronics in Agriculture journal which has maximum citations of 2310 even if this journal has published a smaller number of papers in DL-based leaf disease detection compared to Multimedia Tools and Applications journal, Lecture Notes in Networks and Systems, etc. The sources such as Lecture Notes in Networks and Systems, and Multimedia Tools and Applications journal have started to publish a large number of articles in this domain from 2021. China has published more articles in Frontiers in Plant Science, Computers in Electronics in Agriculture, and Agronomy journal whereas Multimedia Tools and Applications and Lecture Notes in Networks and Systems have more publications in DL-based leaf disease detection from India.

In collaborative research in DL-based leaf disease detection, India holds the top position by collaborating with other 16 countries which produced 725 collaborative publications. Pakistan and South Korea have collaborated with 10 other countries but South Korea has published a smaller number of articles (26) compared to Pakistan (54). Similarly, Egypt and Australia have collaborated with other 8 countries but Egypt published more than double the articles (33) compared to Australia (15).

However, certain limitations were recognized throughout the creation of this study, including the following:

1) While Scopus provides extensive coverage, future studies might consider incorporating data from databases like Web of Science or IEEE Xplore to ensure comprehensive literature coverage.
2) The bibliometric review did not filter journals based on their quality or citation count, resulting in a highly diversified publication.
3) ML-based methods are not considered in the study for leaf disease detection.
4) The author excluded publications written in languages other than English. As a result, relevant research published in various formats and languages may be neglected.
5) Due to low citations in some newly released publications, our bibliometric analysis may not accurately reflect their genuine research rank.

This study's findings highlight potential areas for future investigation which are as follows:

1) Expand the study by searching additional databases for relevant and recent papers.
2) Analyze technique variations over time, and transition from ML to DL in leaf disease detection research.

DECLARATION COMPETING INTERESTS

The authors declare no competing interests.

ACKNOWLEDGMENT

The authors are thankful to the Vellore Institute of Technology (VIT), Vellore for providing all the facilities and support.

REFERENCES

[1] J. Beckman and A. M. Countryman, "The importance of agriculture in the economy: Impacts from COVID-19," *Amer. J. Agricult. Econ.*, vol. 103, no. 5, pp. 1595–1611, Oct. 2021.

[2] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer vision technology in agricultural automation—A review," *Inf. Process. Agricult.*, vol. 7, no. 1, pp. 1–19, 2020.

[3] S. S. Chouhan, U. P. Singh, and S. Jain, "Applications of computer vision in plant pathology: A survey," *Arch. Comput. Methods Eng.*, vol. 27, no. 2, pp. 611–632, Apr. 2020.

[4] R. Sujatha, J. M. Chatterjee, N. Jhanjhi, and S. N. Brohi, "Performance of deep learning vs machine learning in plant leaf disease detection," *Microprocessors Microsyst.*, vol. 80, Feb. 2021, Art. no. 103615.

[5] V. K. Vishnoi, K. Kumar, and B. Kumar, "A comprehensive study of feature extraction techniques for plant leaf disease detection," *Multimedia Tools Appl.*, vol. 81, no. 1, pp. 367–419, Jan. 2022.

[6] M. H. Saleem, J. Potgieter, and K. M. Arif, "Plant disease classification: A comparative evaluation of convolutional neural networks and deep learning optimizers," *Plants*, vol. 9, no. 10, p. 1319, Oct. 2020.

[7] C. Nguyen, V. Sagan, M. Maimaitiyiming, M. Maimaitijiang, S. Bhadra, and M. T. Kwasniewski, "Early detection of plant viral disease using hyperspectral imaging and deep learning," *Sensors*, vol. 21, no. 3, p. 742, Jan. 2021.

[8] L. Li, S. Zhang, and B. Wang, "Plant disease detection and classification by deep learning—A review," *IEEE Access*, vol. 9, pp. 56683–56698, 2021.

[9] S. Vallabhajosyula, V. Sistla, and V. K. K. Kolli, "Transfer learning-based deep ensemble neural network for plant leaf disease detection," *J. Plant Diseases Protection*, vol. 129, no. 3, pp. 545–558, Jun. 2022.

[10] J. Chen, A. Zeb, Y. A. Nanehkaran, and D. Zhang, "Stacking ensemble model of deep learning for plant disease recognition," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 9, pp. 12359–12372, Sep. 2023.

[11] Y. Zhong and M. Zhao, "Research on deep learning in apple leaf disease recognition," *Comput. Electron. Agricult.*, vol. 168, Jan. 2020, Art. no. 105146.

[12] S. M. Hassan, A. K. Maji, M. Jasiński, Z. Leonowicz, and E. Jasińska, "Identification of plant-leaf diseases using CNN and transfer-learning approach," *Electronics*, vol. 10, no. 12, p. 1388, Jun. 2021.

[13] H. Qi, Y. Liang, Q. Ding, and J. Zou, "Automatic identification of peanut-leaf diseases based on stack ensemble," *Appl. Sci.*, vol. 11, no. 4, p. 1950, Feb. 2021.

[14] P. Kokol, H. B. Vošner, and J. Završnik, "Application of bibliometrics in medicine: A historical bibliometrics analysis," *Health Inf. Libraries J.*, vol. 38, no. 2, pp. 125–138, Jun. 2021.

[15] Y. Guo, Z. Hao, S. Zhao, J. Gong, and F. Yang, "Artificial intelligence in health care: Bibliometric analysis," *J. Med. Internet Res.*, vol. 22, no. 7, Jul. 2020, Art. no. e18228.

[16] R. L. Manogna and A. Anand, "A bibliometric analysis on the application of deep learning in finance: Status, development and future directions," *Kybernetes*, vol. 52, no. 10, pp. 1–13, 2023.

[17] G. Sarin, P. Kumar, and M. Mukund, "Text classification using deep learning techniques: A bibliometric analysis and future research directions," *Benchmarking, Int. J.*, Aug. 2023, doi: 10.1108/bij-07-2022-0454.

[18] B. Cibu, C. Delcea, A. Domenteanu, and G. Dumitrescu, "Mapping the evolution of cybernetics: A bibliometric perspective," *Computers*, vol. 12, no. 11, p. 237, Nov. 2023.

[19] Y. Li, Z. Xu, X. Wang, and X. Wang, "A bibliometric analysis on deep learning during 2007–2019," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 12, pp. 2807–2826, Dec. 2020.

[20] A. Abbas, S. Jain, M. Gour, and S. Vankudothu, "Tomato plant disease detection using transfer learning with C-GAN synthetic images," *Comput. Electron. Agricult.*, vol. 187, Aug. 2021, Art. no. 106279.

[21] A. S. Paymode and V. B. Malode, "Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG," *Artif. Intell. Agricult.*, vol. 6, pp. 23–33, Jan. 2022.
[22] R. Karthik, M. Hariharan, S. Anand, P. Mathikshara, A. Johnson, and R. Menaka, "Attention embedded residual CNN for disease detection in tomato leaves," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105933.
[23] G. Shrestha, Deepsikha, M. Das, and N. Dey, "Plant disease detection using CNN," in *Proc. IEEE Appl. Signal Process. Conf. (ASPCON)*, Oct. 2020, pp. 109–113.
[24] D. G. Ranganathan, "A study to find facts behind preprocessing on deep learning algorithms," *J. Innov. Image Process.*, vol. 3, no. 1, pp. 66–74, Apr. 2021.
[25] M. O. Ojo and A. Zahid, "Improving deep learning classifiers performance via preprocessing and class imbalance approaches in a plant disease detection pipeline," *Agronomy*, vol. 13, no. 3, p. 887, Mar. 2023.
[26] A. Fuentes, S. Yoon, S. Kim, and D. Park, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition," *Sensors*, vol. 17, no. 9, p. 2022, Sep. 2017.
[27] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *Eng. Appl. Artif. Intell.*, vol. 115, Oct. 2022, Art. no. 105151.
[28] Y. He, G. Zhang, and Q. Gao, "A novel ensemble learning method for crop leaf disease recognition," *Frontiers Plant Sci.*, vol. 14, Jan. 2024, Art. no. 1280671, doi: 10.3389/fpls.2023.1280671.
[29] P. Kaur, M. P. Singh, A. M. Mishra, A. Shankar, P. Singh, M. Diwakar, and S. R. Nayak, "DELM: Deep ensemble learning model for multiclass classification of super-resolution leaf disease images," *Turkish J. Agricult. Forestry*, vol. 47, no. 5, pp. 727–745, Oct. 2023.
[30] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, p. 6, 2019, doi: 10.1186/s40537-019-0197-0.
[31] M. Li, G. Zhou, A. Chen, J. Yi, C. Lu, M. He, and Y. Hu, "FWDGAN-based data augmentation for tomato leaf disease identification," *Comput. Electron. Agricult.*, vol. 194, Mar. 2022, Art. no. 106779.
[32] J. A. Pandian, G. Geetharamani, and B. Annette, "Data augmentation on plant leaf disease image dataset using image manipulation and deep learning techniques," in *Proc. IEEE 9th Int. Conf. Adv. Comput. (IACC)*, Dec. 2019, pp. 199–204.
[33] A. Pavithra, G. Kalpana, and T. Vigneswaran, "Deep learning-based automated disease detection and classification model for precision agriculture," *Soft Comput.*, pp. 1–12, Mar. 2023, doi: 10.1007/s00500-023-07936-0.
[34] U. Sanath Rao, R. Swathi, V. Sanjana, L. Arpitha, K. Chandrasekhar, Chinmayi, and P. K. Naik, "Deep learning precision farming: Grapes and mango leaf disease detection by transfer learning," *Global Transitions Proc.*, vol. 2, no. 2, pp. 535–544, Nov. 2021.
[35] C. Bajpai, R. Sahu, and K. J. Naik, "Deep learning model for plant-leaf disease detection in precision agriculture," *Int. J. Intell. Syst. Technol. Appl.*, vol. 21, no. 1, p. 72, 2023.
[36] O. Mzoughi and I. Yahiaoui, "Deep learning-based segmentation for disease identification," *Ecological Informat.*, vol. 75, Jul. 2023, Art. no. 102000.
[37] P. Sharma, Y. P. S. Berwal, and W. Ghai, "Performance analysis of deep learning CNN models for disease detection in plants using image segmentation," *Inf. Process. Agricult.*, vol. 7, no. 4, pp. 566–574, Dec. 2020.
[38] L. M. Tassis, J. E. T. de Souza, and R. A. Krohling, "A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images," *Comput. Electron. Agric.*, vol. 186, Jul. 2021, Art. no. 106191.
[39] H.-C. Chen, A. M. Widodo, A. Wisnujati, M. Rahaman, J. C.-W. Lin, L. Chen, and C.-E. Weng, "AlexNet convolutional neural network for disease detection and classification of tomato leaf," *Electronics*, vol. 11, no. 6, p. 951, Mar. 2022.
[40] E. Elfatimi, R. Eryigit, and L. Elfatimi, "Beans leaf diseases classification using MobileNet models," *IEEE Access*, vol. 10, pp. 9471–9482, 2022.
[41] D. Hindarto, "Comparison of detection with transfer learning architecture RestNet18, RestNet50, RestNet101 on corn leaf disease," *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)*, vol. 8, no. 2, pp. 41–48, Nov. 2023.
[42] B. Liu, Y. Zhang, D. He, and Y. Li, "Identification of apple leaf diseases based on deep convolutional neural networks," *Symmetry*, vol. 10, no. 1, p. 11, Dec. 2017.
[43] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," *Comput. Electron. Agricult.*, vol. 173, Jun. 2020, Art. no. 105393.
[44] Z. Qiang, L. He, and F. Dai, "Identification of plant leaf diseases based on inception V3 transfer learning and fine-tuning," in *Communications in Computer and Information Science*. Singapore: Springer, 2019, pp. 118–127.
[45] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," *IEEE Access*, vol. 7, pp. 59069–59080, 2019.
[46] X. Zhang, Y. Qiao, F. Meng, C. Fan, and M. Zhang, "Identification of maize leaf diseases using improved deep convolutional neural networks," *IEEE Access*, vol. 6, pp. 30370–30377, 2018.
[47] G. Geetharamani and A. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," *Comput. Electr. Eng.*, vol. 76, pp. 323–338, Jun. 2019.
[48] A. Islam, R. Islam, S. M. R. Haque, S. M. M. Islam, and M. A. I. Khan, "Rice leaf disease recognition using local threshold based segmentation and deep CNN," *Int. J. Intell. Syst. Appl.*, vol. 13, no. 5, pp. 35–45, Oct. 2021.
[49] M. A. Shahin and S. J. Symons, "Color calibration of scanners for scanner-independent grain grading," *Cereal Chem.*, vol. 80, no. 3, pp. 285–289, May 2003.
[50] N. Marini, S. Otalora, M. Wodzinski, S. Tomassini, A. F. Dragoni, S. Marchand-Maillet, J. P. D. Morales, L. Duran-Lopez, S. Vatrano, H. Müller, and M. Atzori, "Data-driven color augmentation for H&E stained images in computational pathology," *J. Pathol. Informat.*, vol. 14, Jan. 2023, Art. no. 100183.
[51] K. Ma, T. Zhao, K. Zeng, and Z. Wang, "Objective quality assessment for color-to-gray image conversion," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4673–4685, Dec. 2015.
[52] G. Dhingra, V. Kumar, and H. D. Joshi, "Study of digital image processing techniques for leaf disease detection and classification," *Multimedia Tools Appl.*, vol. 77, no. 15, pp. 19951–20000, Aug. 2018.
[53] Z. Zhang, Q. Gao, L. Liu, and Y. He, "A high-quality Rice leaf disease image data augmentation method based on a dual GAN," *IEEE Access*, vol. 11, pp. 21176–21191, 2023.
[54] Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, "Plant leaf disease classification using EfficientNet deep learning model," *Ecolog. Informat.*, vol. 61, Mar. 2021, Art. no. 101182.
[55] *Introduction to Biblioshiny*. Accessed: Jan. 27, 2024. [Online]. Available: https://bibliometrix.org/biblioshiny/biblioshiny1.html
[56] *VOSviewer—Visualizing Scientific Landscapes*. Accessed: Feb. 1, 2024. [Online]. Available: https://www.vosviewer.com/
[57] S. Coulibaly, B. Kamsu-Foguem, D. Kamissoko, and D. Traore, "Deep learning for precision agriculture: A bibliometric analysis," *Intell. Syst. With Appli.*, vol. 16, Nov. 2022, Art. no. 200102, doi: 10.1016/j.iswa.2022.200102.

JYOTISMITA CHAKI (Senior Member, IEEE) received the Ph.D. (Engg.) degree from Jadavpur University, Kolkata, India. She is currently an Associate Professor with the School of Computer Science and Engineering, Vellore Institute of Technology (VIT University), Vellore, India. She has authored and edited many international conferences, journal articles, and books. Her research interests include computer vision and image processing, pattern recognition, medical imaging, soft computing, artificial intelligence, and machine learning. She is an Editor of *Engineering Applications of Artificial Intelligence* (Elsevier); an Associate Editor of *Computers and Electrical Engineering journal* (Elsevier); an Associate Editor of *Array* (Elsevier), *IET Image Processing*, and *Machine Learning with Applications* (Elsevier); and a Section Editor of *PeerJ Computer Science*.

DIBYAJYOTI GHOSH received the Ph.D. degree in industrial engineering and management from IIT (ISM) Dhanbad, in 2018. He holds the position of an Assistant Professor with the VIT Business School, Vellore Institute of Technology, Vellore. He has authored and co-authored various research articles in various peer-reviewed journals and conferences. His primary research interests include supply chain management, artificial intelligence, and optimizations. He is a reviewer for various peer-reviewed journals.