



## Systematic Analysis based on Conflux of Machine Learning and Internet of Things using Bibliometric analysis

Ayushi Chahal<sup>1</sup>, Santosh Reddy Addula<sup>2</sup>, Anurag Jain<sup>3</sup>, Preeti Gulia<sup>4</sup>, Nasib Singh Gill<sup>\*5</sup>, Bala Dhandayuthapani V.<sup>6</sup>

<sup>1,4,5</sup> Department of Computer Science & Applications, Maharshi Dayanand University, Rohtak, Haryana, India.

<sup>2</sup>Department of Information Technology, University of the Cumberland, Williamsburg, KY, USA.

<sup>3</sup>Director, Radharaman Engineering College Bhopal, Madhya Pradesh, India.

<sup>6</sup>Department of IT, College of Computing and Information Sciences, University of Technology and Applied Sciences, Shinas campus, Oman.

Emails: [ayushi.rs.dcsa@mdurohtak.ac.in](mailto:ayushi.rs.dcsa@mdurohtak.ac.in); [santoshaddulait@gmail.com](mailto:santoshaddulait@gmail.com); [anurag.akjain@gmail.com](mailto:anurag.akjain@gmail.com); [preeti@mdurohtak.ac.in](mailto:preeti@mdurohtak.ac.in); [nasib.gill@mdurohtak.ac.in](mailto:nasib.gill@mdurohtak.ac.in); [bala.veerasamy@utas.edu.om](mailto:bala.veerasamy@utas.edu.om)

\*Corresponding Author: Nasib Singh Gill

\*Corresponding Author Email: [nasib.gill@mdurohtak.ac.in](mailto:nasib.gill@mdurohtak.ac.in)

### Abstract

IoT devices produce a gigantic amount of data and it has grown exponentially in previous years. To get insights from this multi-property data, machine learning has proved its worth across the industry. The present paper provides an overview of the variety of data collected through IoT devices. The conflux of machine learning with IoT is also explained using the bibliometric analysis technique. This paper presents a systematic literature review using bibliometric analysis of the data collected from Scopus and WoS. Academic literature for the last six years is used to explore research insights, patterns, and trends in the field of IoT using machine learning. This study analyses and assesses research for the last six years using machine learning in seven IoT domains like Healthcare, Smart City, Energy systems, Industrial IoT, Security, Climate, and Agriculture. The author's and country-wise citation analysis is also presented in this study. VOSviewer version 1.6.18 is used to provide a graphical representation of author citation analysis. This study may be quite helpful for researchers and practitioners to develop a blueprint of machine learning techniques in various IoT domains.

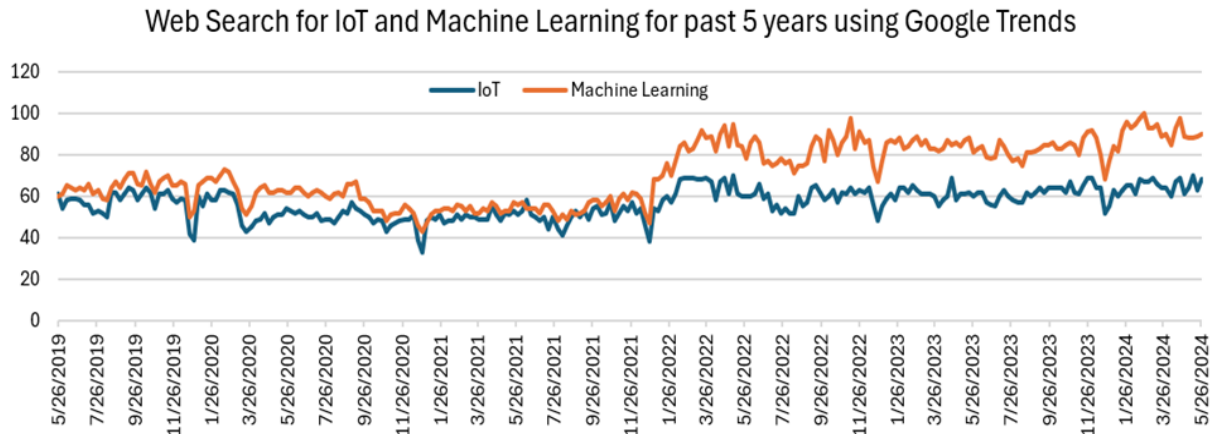
Received: September 29, 2023 Revised: January 27, 2024 Accepted: June 11, 2024

**Keywords:** IoT; Machine Learning; Citation Analysis; Healthcare; Smart City; Security; IIoT.

### 1. Introduction

With the rapid advancement in hardware and software to collect data, it is needed to have different techniques or methods to handle that data. But when communication technologies like the Internet get connected with different sensory devices then digital data from the physical world is collected using the Internet of Things. It has increased the amount to collect data exponentially. Now-a-days, IoT has become an integral part of almost every progressive sector of society like smart homes, smart healthcare systems [1], smart education sectors, smart agriculture, smart industries, smart wearable things, etc. IoT helps to connect these devices through a wireless medium. The Internet helps to collect a large amount of heterogeneous data from every part of the world.

Data generated by IoT devices like sensors, actuators, etc. is heterogeneous, dynamic, and uncertain in nature. This huge wealth of data collected by IoT devices needs to be analyzed. Such a large amount of data collected



by IoT devices is not possible for humans to explore by themselves. To get insights from collected data, artificial intelligence techniques Machine Learning (ML), and Deep Learning (DL) play a very important role. Deep learning is considered as a part of machine learning only. Various machine learning approaches are used in every sector of daily routine life like education, banking, healthcare, security, finance, etc. directly or indirectly with respect to intelligent systems [2]. Figure 1 displays the trend of web search on the Google search engine for machine learning and IoT for the past 5 years.

Figure 1: Web search for machine learning and IoT for past 5 years using Google Trends.

The field of IoT and Machine learning is very vast and popular among researchers now-a-days. Before moving to any new concept, researchers have to go through a large academic literature to get insights. It is difficult to read thousands of papers and extract useful information from those papers. Bibliometric analysis is a technique that eases this difficulty in an efficient manner. The bibliometric analysis tends to answer some research-oriented questions using a large number of publications [3].

A Systematic Literature Review (SLR) guideline helps to cover a large database of articles and pick those articles which are suitable for one's research purpose. A. Ligthart et al. [4] presented a tertiary study on sentimental analysis based on SLR principles which are found to be very beneficial to provide a correct direction to the research. SLR principles are used to make a quick study about the readability of structured and unstructured abstracts. By following SLR steps they achieved their goal. Making IoT systems intelligent, with the use of machine learning, is a huge and dynamic platform. The number of research and review articles gets updated on various platforms on daily basis. Such a large number of articles are not easy to review without any guidance, hence Systematic Literature Review (SLR) comes into play. SLR rules will help us to filter out important articles which will help us with further research.

The main objectives of this study are:

- This study provides a detailed overview of different kinds of data generated using IoT sensor devices.
- This study provides a systematic literature review of machine learning techniques that can be used for IoT data analysis in various domains.
- Superlative literature is identified for the last 6-year academic research dataset using the Bibliometric analytical technique, which helps to find influential authors of respective domains.
- This study helps to understand the document-citation network and co-authorship among influential authors in the Machine Learning and IoT field.

### A. IoT Data

IoT is a data-collecting body that uses different sensory devices and the internet to collect the data. IoT devices work continuously to grab the data and as a result generated data is of huge amount which cannot be handled by a single system. This data needs to be handled cautiously so that the data can be converted from information to fruitful insights. For this purpose, different data analysis steps are applied to this collected data. Before moving to analysis part, let's just have look on heterogeneous IoT data than needed to be handled with care. This study divides IoT data types in two parts i.e., Primary data and Secondary data as shown in Figure 2.

a. Primary Data

Data collected by IoT devices are heterogeneous in nature on the basis of their structure. Primary data is categorized into three forms of data types based on the storage structure. These are structured, semi-structured, and unstructured.

- Structured Data

These data can be structured, semi-structured, or unstructured. Data that can be stored in some format like tabular format etc. is called structured data. Relation datasets that have tuples and attributes stored in the form of a table are a perfect example of structured data.

- Semi-structured Data

Semi-structured data does not have any structure, but it has a certain kind of layout to store data in a certain way. Java Script Object Notation (JSON) files, XML files, and graph databases are some examples of semi-structured data.

- Unstructured Data

Unstructured data is data which do not have any kind of pre-defined structure or layout. Video files, audio files, and heavy binary data files are some examples of unstructured data. Most of the data a researcher gets from IoT devices for data analysis is unstructured data.

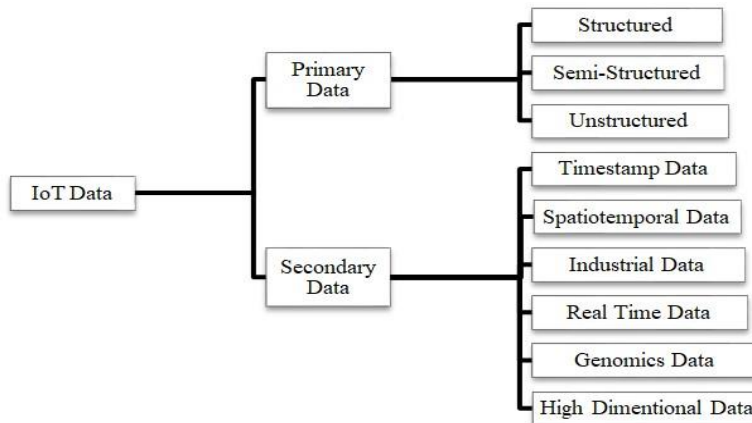


Figure 2: Variety of IoT Data

b. Secondary Data

Secondary data categorize data on the basis of its usability of the data while primary IoT data is the basic nature of the data being used. When primary data is associated with certain add-on features, they generated secondary data. This study describes the following type of secondary data types:

- Timestamp Data

Data that contains two values attached with it i.e. time and date is regarded as Timestamp data. IoT devices are able to collect data from the surrounding environment but not necessarily all the time. To save some of the bandwidth data collected by IoT devices are for some particular instances depending on their utilities. To know the exact timing and day of the data taken (time, date) is necessary to insert in the datasets [5].

- Spatiotemporal Data

Data that has a location (space) and time as its key features are known to be spatiotemporal datasets. Many IoT sensor devices are attached to moving objects. These moving objects change location with the change in time. A lot of data that are used for the analysis of traffic flow, weather prediction etc. are examples of this kind of dataset. Analysis of spatiotemporal data is considered a bit difficult because both attributes location and time vary in every second but it is still possible by various machine learning algorithms.

- Industrial Data

Industries like electricity generating industries, chemical industries, oil-petroleum industries, and manufacturing industries uses boiler feeds, turbines, pumps of different size. These industries collect data like pressure, humidity, temperature, etc. with the help of IoT sensors. This data is used for analysis like predicting pump failure, calculating the life of a turbine, etc. Data used in these industries are called industrial

data. Machine Learning algorithms help in analyzing and predicting threats as well as troubleshooting the problems in that industry.

- **Real-Time Data**

Data which is directly transferred from an IoT device to the end user for analysis, without being stored anywhere is called real-time data. This huge amount of real-time data collected by IoT sensor devices is handled with care and can be used for predictive analysis with the help of machine learning algorithms. Such kind of data is used in hospitals to send pulse rates on the screen, in driverless cars to identify risks nearby, etc.

- **Genomics Data**

This data is basically related to healthcare industries. To predict disease and analyze neurological characteristics of the patients, many IoT wearable devices are used like Fitbit One etc. These devices can also be used to collect biospecimen. This biospecimen can help medical scientists to analyze DNA of the person with the help of different machine learning algorithm. After analysis, they may provide any kind of customized treatment to the person according to its DNA properties. This analysis can also help in the discovery of new drugs. A complete process of collecting biospecimens of a person with the help of IoT wearable devices is explained in the study.

- **High Dimensional Data**

Different IoT sensor devices collect high-dimensional data from the surroundings. Data is considered to be high dimensional when number of attributes of data are equivalent or may outstrip the number of observations. Before applying any kind of data analysis this data, needs to be reduced. Here comes the part of Dimensionality Reduction. Dimensionality Reduction helps in reducing the number of features with respect to observations. Different machine learning algorithms like support vector machine (SVM), decision tree, or regression can further be used for data analysis.

### B. Conflux of Machines Learning and IoT

IoT has become an integral part of day-to-day life. IoT provides Machine-to-Machine communication [6]. It has made human life very easier by spreading its roots to almost every sector that is related to humans. Different devices like smart watches, oxi-meters, smartphones, and fitness bands are the most basic gadgets that almost every person carries with them. Similar to these gadgets, there are multiple tools used in various companies and organizations at large scale.

These gadgets generate a lot of data. IoT data is heterogeneous, dynamic, uncertain, unbalanced, and large in volume. Data with so many random properties is needed to be analyzed to reach any conclusion from the data. These random properties of a dataset can be handled by artificial intelligence techniques, specifically machine learning. Machine learning algorithms outperform other technologies in the field of data analysis. Data analysis can be of many forms i.e. Descriptive, Predictive, Perspective, and Adaptive.

Machine learning is the most efficient data analysis technique that can be used to analyze IoT data. Different machine learning techniques are being improved according to the change in the variety of the data collected by the IoT devices. Machine learning techniques are divided into four categories i.e. Supervised, Semi-supervised, Unsupervised, and Reinforcement. Figure 3 presents different categories of machine learning techniques with respect to data types.

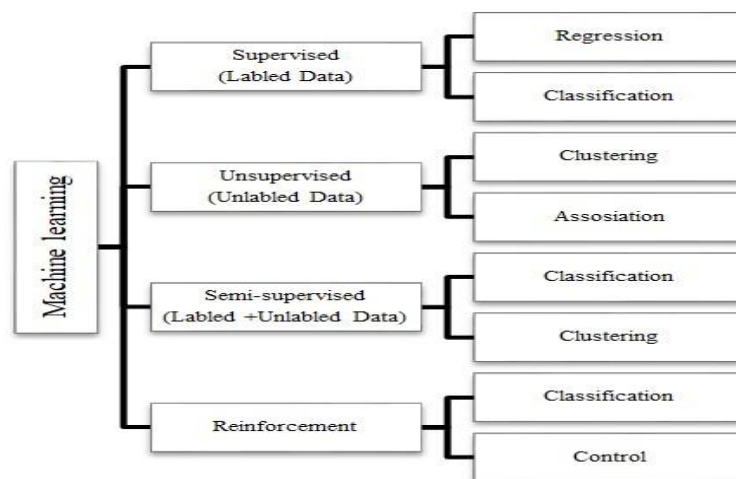


Figure 3: Machine Learning techniques categorization with respect to data types

- Supervised learning algorithms can be applied to those data which are labeled, or which have some predefined outcomes. Two kinds of supervised techniques can be used to analyze the data i.e. Regression and Classification. Many data sets of the medical field for disease detection, are labeled are examples of labeled data.
- Unsupervised learning algorithms can be applied to those datasets which are not labeled. These algorithms are generally used to identify some patterns hidden in the data. Clustering is a widely accepted unsupervised machine learning technique.
- Semi-supervised learning algorithm is applied to the data that is a mixture of labelled and unlabelled data. One generally tackles these kinds of data in real life because labeled data is less available than unlabelled data.
- Reinforcement learning algorithm is applied to action-oriented data. It helps in learning through hit-and-trial methods. In this kind of learning approach, actions are taken according to the observations in order to maximize the output and minimize the risk factor.

IoT and machine learning go hand in hand in the analysis of data. Table 1 explains the benefits of machine learning for an IoT-enabled environment based on seven selected application sectors. These application sectors are depicted in Figure 4 and described below:

Table 1: Main benefits of machine learning in IoT

S.NO.	Benefits
1.	Enhance security, threat & intrusion detection over IoT networks
2.	Pest and disease detection, classification of crop, Yield estimation, Water requirements, soil moisture, etc.
3.	Parking lot and traffic congestion prediction, road surveillance and smart home facilities etc.
4.	Electric Grid load prediction, load balancing, smart meters, etc.
5.	Disease detection & prediction, wireless healthcare monitoring system, etc.
6.	Weather forecasting, rainfall prediction, humidity prediction, etc.
7.	Manufacturing machine's life prediction, Robot domination Industries, etc.

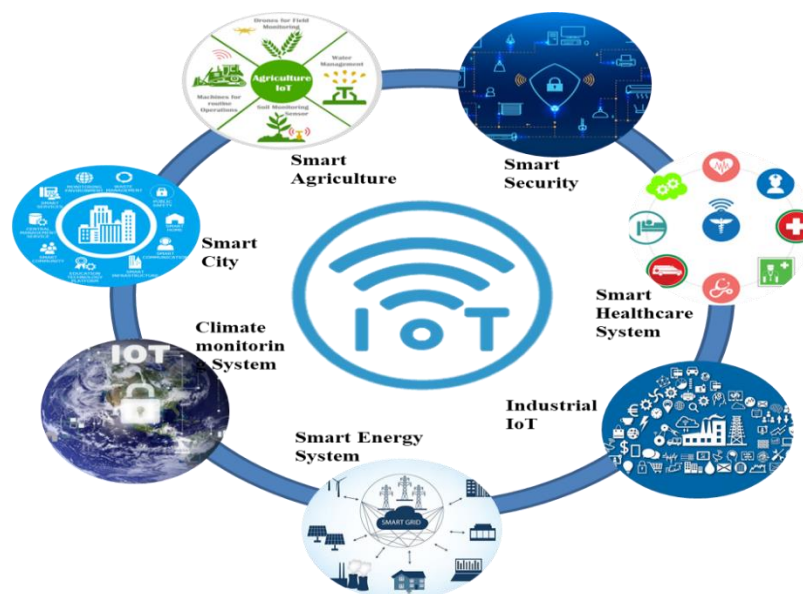


Figure 4: Machine learning contributing to various IoT domains

#### a. Healthcare

The Healthcare Sector is a rapidly developing sector in IoT. The amount of data collected in this sector is enormous due to wearable sensor devices, and health monitoring systems of a patient. Most of the time data generated in this sector are unstructured data. Data collected in the healthcare sector are from X-rays, CT scans, EEG, ECG, blood

pressure, body temperature, sugar level, respiratory rate etc. Moreover, patients in ICUs that require continuous monitoring have sensors all over the body. These sensors generate continuous data. To analyze such a large amount of data and predict the results machine learning algorithms are used. Mostly used machine learning algorithms for this sector are logistic regression, SVM, and J48 classifier.

#### b. Agriculture

The economies of many countries like India are based on agriculture. Smart farming and smart irrigation are some of the new techniques that help in increasing production. To automate farming or forestry and include intelligence in it IoT techniques are needed to be implemented. Also, machine learning and cloud computing can help in automating farming. IoT has different sensory devices that can be used for measuring soil moisture, humidity in the environment, soil temperature, ultraviolet radiations, and many more [7].

#### c. Climate

Climate monitoring and weather forecasting systems are already present in the market. These systems work in collaboration with different IoT sensors and for prediction, it use machine learning and deep learning technology. Different kinds of IoT sensors that can be used for climate monitoring are temperature sensors, humidity sensors, air flow sensors, light sensors, etc.

#### d. Energy

With the increase in the development of Smart City projects, conservation of energy resources and management of energy systems become the main concern. To save energy-producing resources, the combination of IoT tools and machine learning techniques is quite helpful. For smart city development, the smart grid plays an important role. Smart grids help in maintaining energy release. It manages energy generated, energy used, energy remained to perform any task. [8]

#### e. Industrial IoT (IIoT)

Today's era of industries is trying to automate every machine. Smart devices used in industries perform their functioning using machine learning concepts. Industry 4.0 is considered the fourth revolution in industries. Smart manufacturing industries are recent trends in the industrial world. The connection between different machines and devices in industries is set up with the help of the internet which takes a step toward the next revolution of industries. This upgradation of industries is called the Industrial Internet of Things (IIoT). Machine learning concepts are used to monitor and automate these IIoT machinery. IIoT basically uses the concept of IoT i.e. to connect anything, anywhere, anytime.

#### f. Security

The security of data is always a concern for any technology. IoT generates thousands of gigabytes of data that contain some critical information like personal data, confidential data, etc. which need to be protected. To avoid any kind of security breach, many machine learning algorithms like decision trees, Naive Bayes (NB), and Artificial Neural Networks (ANN) collaborate with an Intrusion Detection System (IDS).

#### g. Smart City

To make personal life easier, cities are also smart now-a-days. A smart city includes lot of larger concepts within it like a smart home, smart transportation system, smart parking system, etc. Different home appliances like washing machines, televisions, air conditioners, etc. can be automated using IoT and machine learning. Smart transportation includes smart parking systems, road surveillance, traffic prediction, etc. All these systems make the life of a common man much easier and simplified.

## 2. Methodology

This study is based on a bibliometric analysis of machine learning algorithms used for the analysis of data collected by IoT sensor devices. The bibliometric analysis technique is used to find insights and summarize existing literature using quantitative methodology. Generally, Bibliometric analysis uses a large dataset to get insights [9]. A systematic literature review is conducted on different published articles using methods shown in Figure 5 explained by Charter et al. in their study [10]. Before these methodological steps, data collection and citation analysis are required.

### A. Data Collection

To cover maximum research literature data is collected from Scopus and Web of Sciences. Meta-data related to the publications for Machine learning, IoT, and the respective domain is exported from these well-known platforms. This dataset is further used for this study. Firstly, duplicate data are removed from the dataset. Then, literature data is analyzed, and commonly used keywords are recognized. Document citation analysis is then used to identify the relationship between the document and the authors. It helps to find the most cited paper and most popular author for a particular field. These connections and patterns help to cluster the publications on the basis of

similarity. Using the bibliometric analysis technique on this meta-data, the contribution of different countries in a particular research field is also studied.

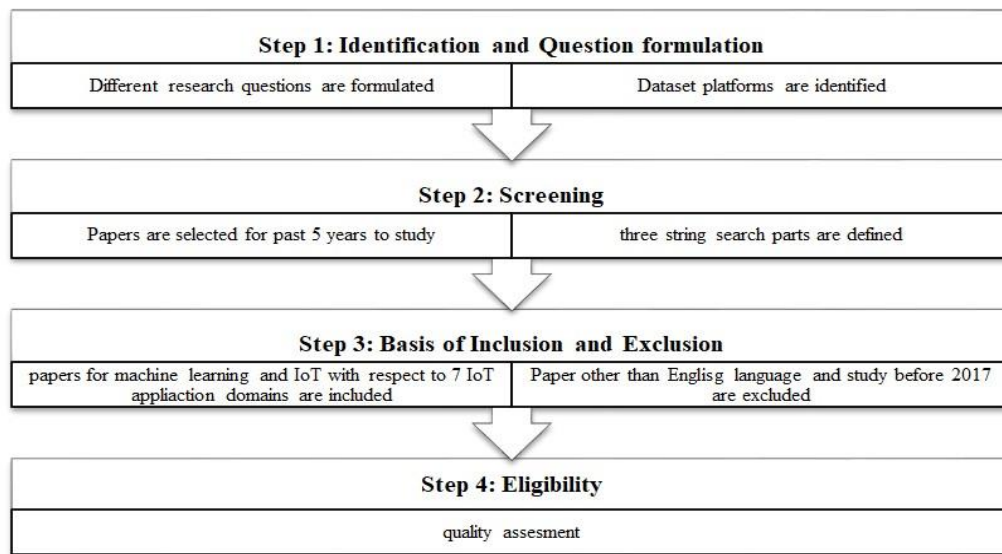


Figure 5: Complete methodology for the review process

## B. Citation Analysis

Insights of the research literature are revealed using Citation analysis. Citation analysis helps to identify the relativity among the items used for analysis, on the basis of how these items cite each other. Items used for citation analysis are documents, sources, organizations, authors, and counties. Citation analysis can be done by various means such as document citation analysis, author citation analysis, organizations included in the citation, country-wise citation, and sources of citation. Every kind of citation analysis plays a vital role in enhancing the quality of bibliometric analysis. Citation analysis helps to find the most influential publications and authors in a respective research field. Citation analysis also helps to track the evaluation of the knowledge in a particular research area over a period of time.

## C. Methodology Steps

To improve the standard of the paper bibliometric analysis is done using an updated version of the systematic literature review.

- Step1: Identification and question formulation

This step deals with the search strategies for the review process. There is a vast variety of literature present on various platforms. Data is gathered from different databases like Web of Science (WOS), and Scopus. In this step, research questions are formulated that are to be tackled like:

- a. What is the state-of-art approach for using different machine learning techniques for IoT data analysis?
- b. What are the research trends in the literature of machine learning with respect to IoT data used in various IoT fields?

- Step 2: Screening

This step deals with the selection criterion for the review process. All the articles selected for the review process are confined to the period of 2017 to 2022. For the applicability of these screening criteria to every IoT sector, the search string is divided into three parts to extract data from databases of every platform. The first part is “machine learning”, which is used to specify the usage of machine learning techniques. The second part is “IoT”, which is used to ensure it can relate to both machine learning and IoT sectors. The third part depends upon different IoT application sectors. This interrelation between search strings helps to find the majority of machine learning techniques used in different IoT application sectors.

- Step 3: Basis of Inclusion and Exclusion

This step helps extract data from the complete dataset that is useable for the study. All the duplicate data are removed from the extracted data. To refine the study results inclusion and exclusion criteria needed to be defined. Any feature defined in the inclusion category needs to be included in the study and any feature defined in the exclusion category needs to be strictly excluded from the study.

- a. For this paper inclusion criteria are:
  - (1) This study includes Systematic Literature review studies.
  - (2) This study is strictly related to machine learning and IoT.
  - (3) This study includes papers from different IoT application sectors.
- b. For this paper exclusion criterion is:
  - (1) To study except the English language is not allowed.
  - (2) Studies before 2017 are not allowed to be part of the systematic review process.

- Step 4: Eligibility

This step deals with the quality assessment for the review process. This step is done manually. After reading multiple extracted papers, the outcomes are matched with the requirements. If they matched, then the SLR process is considered as successful.

### 3. Results

Results are found with respect to various application sectors of IoT where machine learning algorithms are applicable for analysis purpose. Web of Sciences and Scopus datasets are explored to extract multiple papers. This section gives a sector-wise brief overview to an in-depth analysis of many pertinent papers which are helpful to get the results.

#### A. Healthcare Sector

For analysis purpose, strings used for searching publications and outcomes for every search are shown in Table 2.

Table 2: Different datasets used for analysis

S. No.	Data Set	Keyword set for searching	#Papers found
1.	WOS	TOPIC: (Machine Learning) AND TOPIC: (IoT) AND TOPIC: (healthcare) Timespan: 2017-2022.	401
2.	Scopus	Title, abstract, keywords: ‘Machine Learning’ AND ‘IoT’ AND ‘Healthcare’	1078

- a. Bibliometric Analysis w.r.t. Scopus

For Bibliometric analysis, only Web of Sciences and Scopus data have been used in this study. After the removal of duplicates in Scopus data, 1078 publications have been found. 96 documents are filtered out on the basis of most cited (more than 25 citations) documents which are clustered on the basis of their connectivity strength. In this analysis, 81 such types of clustered are formed. Using these clusters, most cited papers are found and further used to survey different machine learning algorithms applicable to the IoT healthcare sector. Figure 6 represents the document citation analysis.

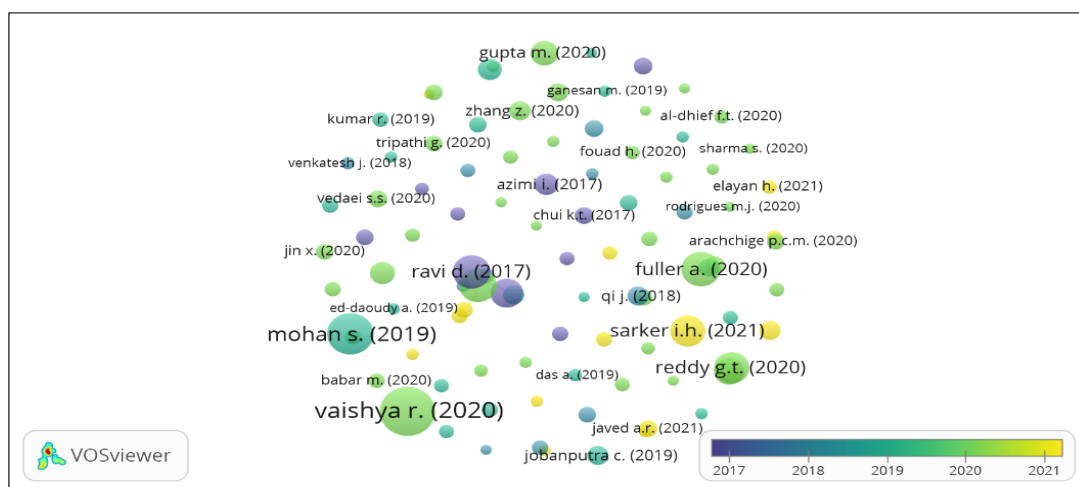


Figure 6: Document citation analysis using Scopus data for the Healthcare sector

As Figure 6 shows, the study by R. Vaishya et. al (2020) is the most cited document with 532 citations. R. Vaishya et. al (2020) discussed the usage of Artificial Intelligence (AI) or Machine Learning (ML) for the development of

drugs and vaccines for the current pandemic COVID-19. The authors explained how IoT becomes helpful for this pandemic scenario. IoT has become helpful in this pandemic scenario. J.P.O. Li et. al (2021) presented a study on telemedicine and AI techniques that are popular in the field of ophthalmology now-a-days [11]. A.R. Javed et. al (2021) have used the Ada-Boost machine learning algorithm to detect cognitive impairment among individuals in their early stages [12].

Supervised, unsupervised, and semi-supervised machine learning techniques in IoT-enabled healthcare monitoring systems can be used for disease diagnosis systems [12]. Optimized machine learning algorithms can help in analysis of the data collected through IoT sensors for cardiovascular disease, diabetes, Alzheimer’s, tuberculosis, etc. [13]. Multiple machine learning algorithms like Decision Trees, Support Vector Machine (SVM), Naive Bayes Classifiers, and Random Forest Classifiers can be applied on any kind of high-dimensional IoT data like healthcare CTG cardiography, etc. for dimensionality reduction. Machine learning algorithms like the Random Forest, KNN, Decision Tree, etc. are used in IoT-based health monitoring systems [15]. The Internet of Robotic Things (IoRT) uses technologies like machine learning, IoT, cloud computing, etc. to work in the healthcare sector. It helps to insert small robotic inhalable devices into the patients to get an idea of organs from inside without any kind of surgery.

Country-wise citation analysis in the field of IoT Healthcare using machine learning is shown in Figure 7. The threshold for this study has been set where the minimum no. of documents of countries should be 12 and the minimum no. of citations for a country should be 98. Using this threshold, 25 countries are filtered out which are further clustered into 4 clusters. Among all resulted 25 countries, India is at the top with 3916 citations. The top 5 countries with the number of publications and total citations are shown in Table 3.

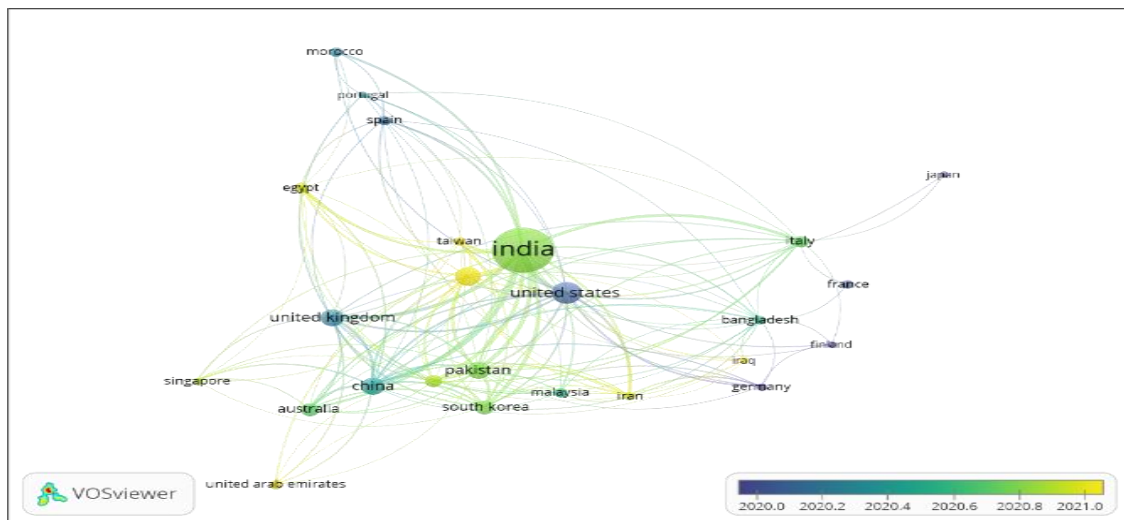


Figure 7: Country-wise citation analysis using Scopus data for the Healthcare sector

Table 3: Country-wise citation analysis on Scopus database for the healthcare sector

Country	#Documents	#Citations
India	460	3916
China	66	1780
United Kingdom	68	1733
United States	110	1663
Pakistan	68	975

b. Bibliometric Analysis w.r.t. Web of Science

For bibliometric analysis from the WoS database, 401 studies have been extracted. These studies have no duplicate records. 1714 authors contributed a total of 401 publications in the research fields of Smart Healthcare using IoT and machine learning. Co-author citation analysis is done by applying a threshold on several documents and a number of citations. To get the results, the minimum number of documents per author is set to 3 and the minimum number of citations for authors is set to 15. Based on this criterion, the highly cited work of 36 authors is considered for this study. Figure 8 presents the co-author citation analysis results.

Some of the studies are indexed by both Scopus as well as WoS [14]. A detailed review is presented by H.K. Bharadwaj et. al (2021) [16], for machine learning algorithms to enable IoT healthcare applications. A new term H-IoT is introduced. This study has covered all the steps used for H-IoT like diagnosis, prognosis and spread

control, assistive systems, monitoring, and logistics. Machine learning techniques can be used for prediction model with its classification properties.

Figure 9 shows the most contributed countries in the field of IoT Healthcare and used machine learning in their research. The threshold is set where the minimum no. of documents of countries is 12 and the minimum no. of citations for a country is 100. Using this threshold, 24 countries are filtered out from a total of 80 countries. These are further clustered into 4 clusters. Among all resulted 25 countries, India is at the top with 1717 citations. The top 5 countries with the number of documents published and citations for this sector are given in Table 4.

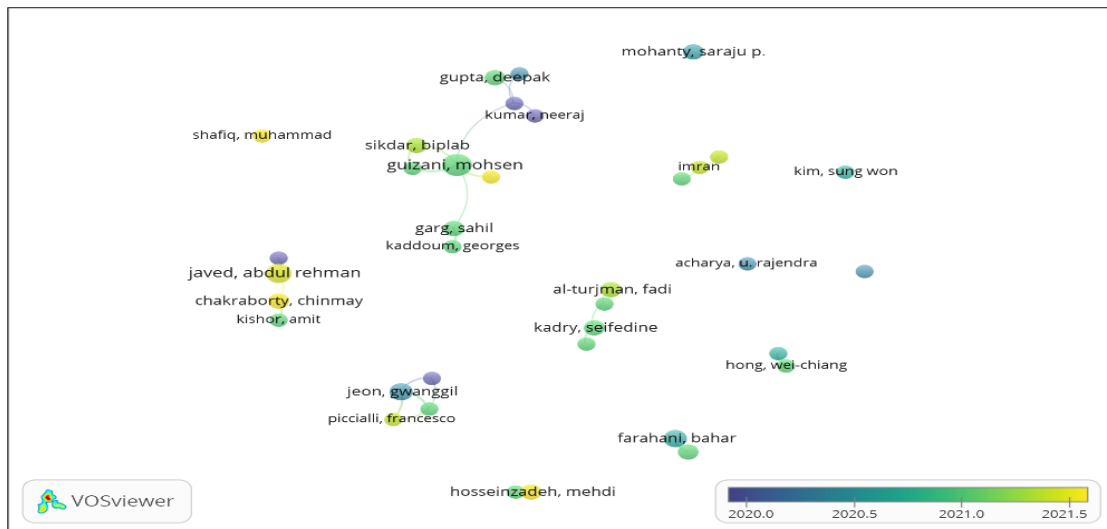


Figure 8: Co-Author citation analysis using WoS data the Healthcare sector

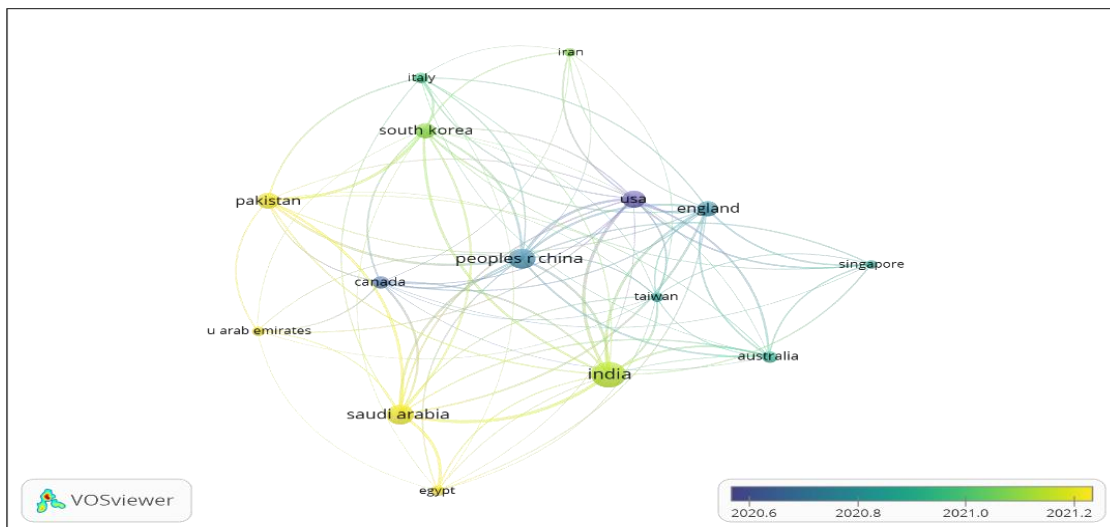


Figure 9: Country-wise citation analysis using WoS data for the Healthcare sector

Table 4: Country-wise citation analysis on WoS database for the Healthcare sector

Country	#Documents	#Citations
India	126	1717
Saudi Arabia	75	1500
China	43	992
USA	53	955
South Korea	46	613

## B. Agriculture

For analysis purpose, strings used for searching publications and outcomes for every search is shown in Table 5.

Table 5: Different datasets used for bibliometric analysis

S. No.	Data Set	Keyword set for searching	#Papers found
1.	WOS	TOPIC: (Machine Learning) AND TOPIC: (IoT) AND TOPIC: (Agriculture) Timespan: 2017-2022.	227
2.	Scopus	Title, abstract, keywords: ‘Machine Learning’ AND ‘IoT’ AND ‘Agriculture’	615

### a. Bibliometric Analysis w.r.t Scopus dataset

For bibliometric analysis, 612 studies have been extracted from the Scopus database. These studies have no duplicates. 67 documents have been filtered out on the basis of most cited (more than 15 citations) documents, which are clustered in 55 clusters on the basis of their connectivity strength. Figure 10 presents the document citation analysis results for Scopus data.

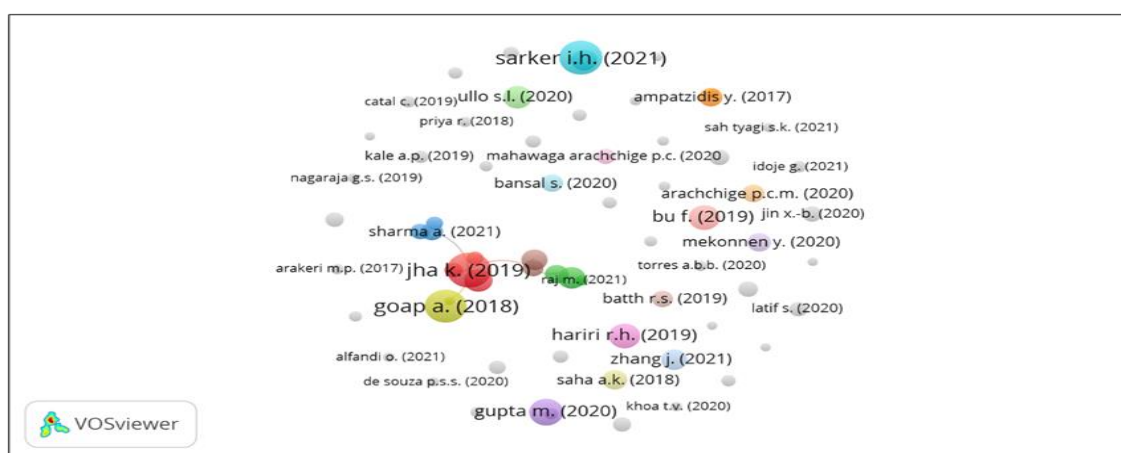


Figure 10: Document citation analysis using Scopus data for the agriculture sector

K. Jha et. al (2019) [17], is found to be the most cited document among all the 67 extracted documents. This study has 243 citations. The authors have explained different wireless systems to be used in agriculture. They employed IoT and deep learning techniques to classify different plants and flowers. The importance of artificial intelligence techniques is described for sustainable farming. Smart irrigation system uses different sensors for data acquisition like soil moisture, temperature, UV radiations, air temperature, etc. Machine learning algorithms can work on weather data to predict the future environment. Generally, temporal datasets are used to forecast crop harvests [18].

Country-wise citation analysis is done over the same data collected from Scopus. To get results, the threshold for a minimum number of documents for a country is set as 5 and the minimum number of citations per document selected is set as 15. 27 countries out of 96 are selected with respect to this threshold value, among which India is at the top with 2090 citations. Figure 11 presents country-wise citation results. The top 5 countries with a number of documents published and total citations are shown in Table 6.

Table 6: Country-wise citation analysis on Scopus database for the agriculture sector

Country	#Documents	#Citations
India	294	2090
United States	45	766
Australia	17	570
China	28	243
Italy	20	295

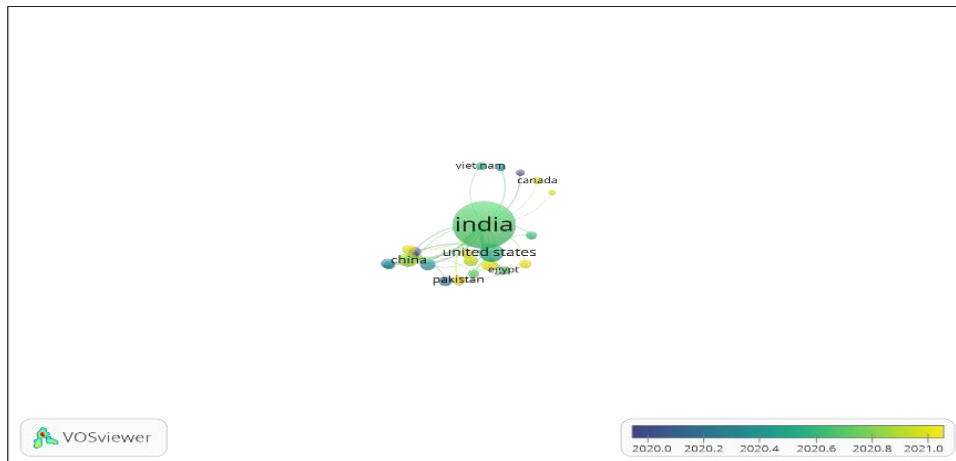


Figure 11: Country-wise citation analysis using Scopus data for the agriculture sector

b. Bibliometric Analysis w.r.t. Web of Science

For bibliometric analysis, 227 studies have been extracted from the WoS database. These studies have no duplicates. 10 documents are filtered out on the basis of the most cited (more than 15 citations) documents, which are clustered into 6 clusters on the basis of their connectivity strength. Figure 12 presents the connections among the authors of the documents cited.

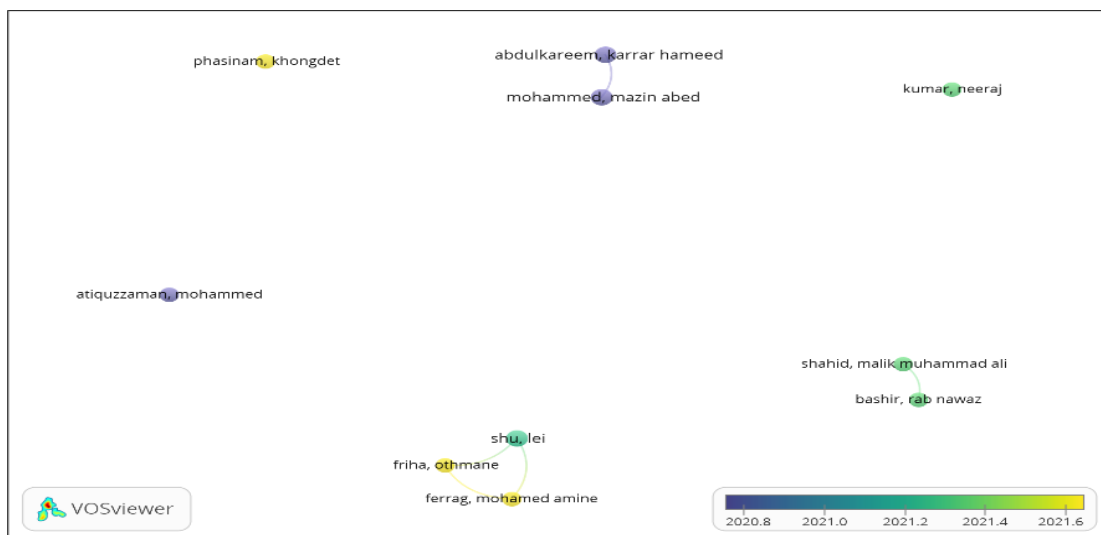


Figure 12: Document citation analysis using WoS data for the agriculture sector

Some of the studies are indexed in the database of Scopus as well as WoS. IoT plays an important role in ‘arable’ farming [19]. The authors have discussed in-depth implementation and application of IoT in farming. To overcome the challenges during the implementation of IoT in smart agriculture, techniques like machine learning, middleware platforms, etc. play an important role.

Country-wise citation analysis is done over the same data collected from WoS. To get results, the threshold for a minimum number of documents for a country is set as 5 and a minimum number of citations per document selected is set as 15. 16 countries out of 69 are chosen with respect to this threshold value. Figure 13 presents country-wise citation results, among which India is at the top with 689 citations for 69 documents. The top 5 countries with several documents published and their total citation are shown in Table 7.

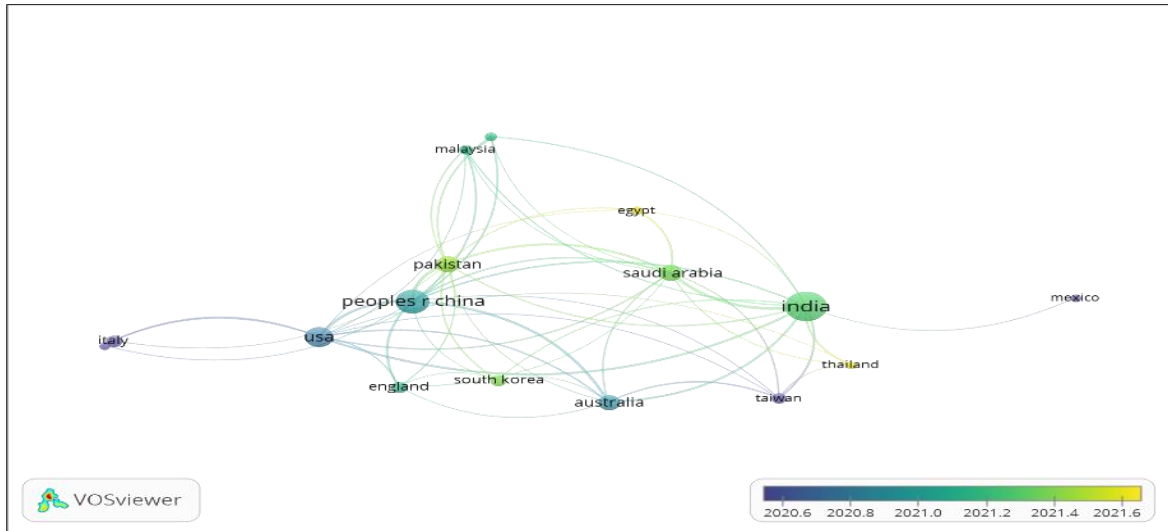


Figure 13: Country-wise citation analysis using WoS data for agriculture sector

Table 7: Country-wise citation analysis on Scopus database for agriculture sector

Country	#Documents	#Citations
India	69	689
United States	33	625
China	47	332
Australia	19	322
Italy	11	226

### C. Climate Sector

For analysis purpose, strings used for searching publications and number of outcomes for every search is shown in Table 8.

Table 8: Different datasets used for analysis

S. No.	Data Set	Keyword set for searching	#Papers found
1.	WOS	TOPIC: (Machine Learning) AND TOPIC: (IoT) AND TOPIC: (Climate) Timespan: 2017-2022.	60
2.	Scopus	Title, abstract, keywords: 'Machine Learning' AND "IoT" AND "Climate"	154

#### a. Bibliometric Analysis w.r.t Scopus dataset

For bibliometric analysis, from the Scopus database, 154 studies have been extracted. 14 documents have been filtered out of 154 on the basis of most cited (more than 15 citations). These 14 highly cited studies have been depicted in Figure 14.

J. Pang et. al, 2021 [20] is found to be most cited publication among other extracted studies after applying the threshold on it. This study scored 43 citations as of now. It considers the heterogeneity of the data to develop a self-organized Federated Learning framework. This framework used reinforcement learning methods. IoT sensor can be used to check indoor air quality. IoT sensor devices acquire data from air particles such as content of CO<sub>2</sub>, NO<sub>2</sub>, humidity, CO content in the air and then use machine learning technologies to analyze whether this is good for health or not. Result of the study concluded that larger the number of people in the house, lower will be the air quality and more the ventilation in the house, better will be the air quality [19]. Supervised machine learning modules are used to predict climate conditions. Temperature sensors can be used for data acquisition which collects spatiotemporal data. A hydrocarbon network model helps to predict the temperature from remote locations also [21]. J.J. Aguilera et. al (2019) [22], created Personal Comfort Model which computes individual thermal need with the help of IoT thermal sensor devices. It uses machine learning techniques to predict and provide the exact amount of heat and air quality that is needed indoors. M. Alhussein et. al (2019) [23], proposed an energy management system that is based on Distribute Energy Resources (DER). Authors have used deep learning algorithms to predict wind speed and solar irradiance. This proposed model helps people to reduce electricity bills, by predicting energy generated by DERs.

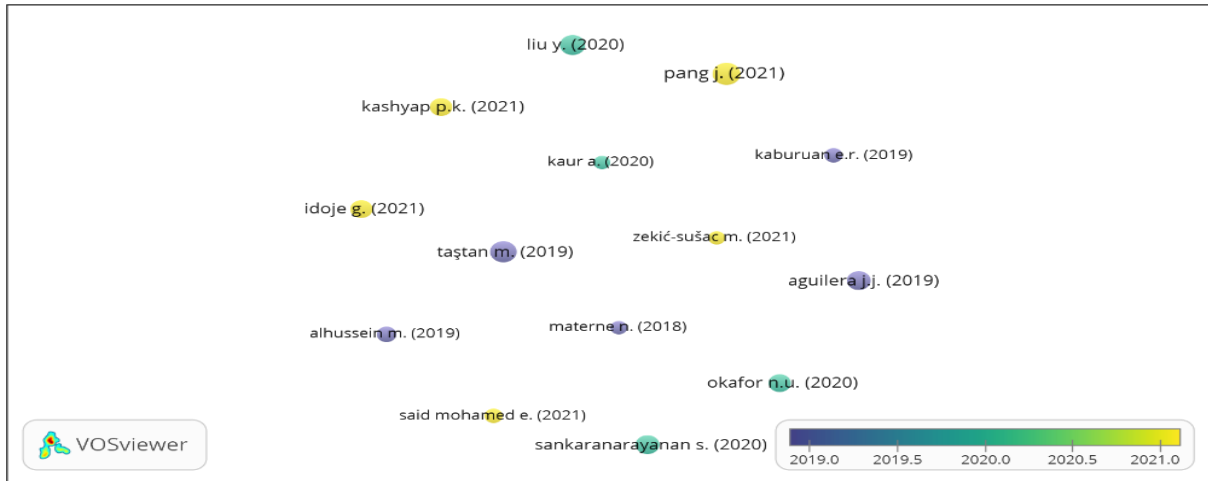


Figure 14: Document citation analysis using Scopus data for the Climate sector

Country-wise citation analysis is done using the Scopus dataset. The threshold is maintained for this analysis as the minimum number of authors for a publication is set as 5, and the minimum number of citations with respect to a country must be at least 15. 8 out of 52 countries came as a result with respect to the threshold value. Among all these countries, India again is at the top with 169 citations. Figure 15 presents the country-wise analysis plot using VOSviewer. The top 5 countries with a number of documents and total citations are shown in Table 9.

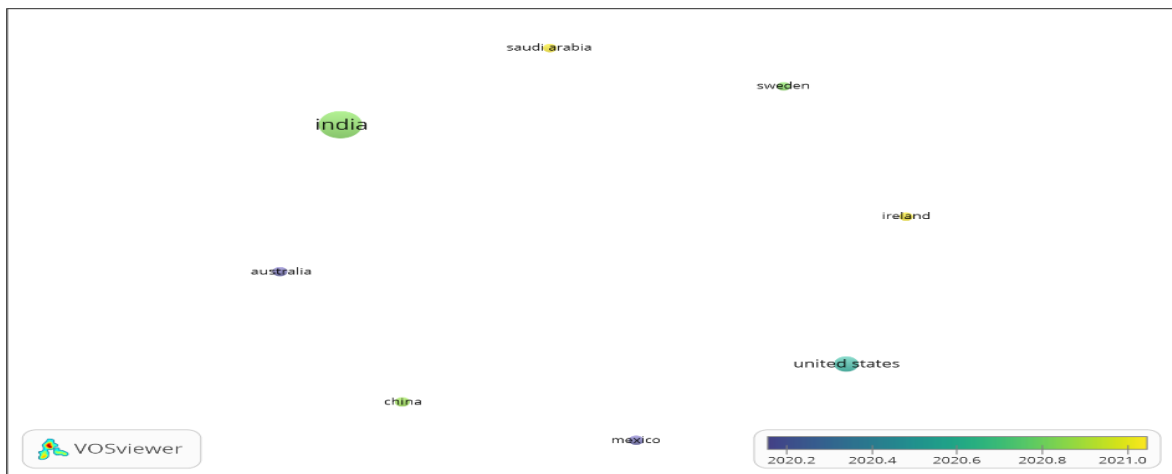


Figure 15: Country-wise citation analysis using Scopus data for the Climate Sector

Table 9: Country-wise citation analysis on Scopus database for Climate Sector

Country	#Documents	#Citations
India	51	169
United States	16	100
Ireland	5	64
China	6	50
Sweden	5	39

b. Bibliometric Analysis w.r.t. Web of Science

For bibliometric analysis from the WoS database, 60 studies have been extracted. These studies have no duplicate records. 3 documents have been filtered out of 60 on the basis of most cited (more than 3 citations). Some studies like [21, 23] are indexed by both Scopus and WoS. Personal Comfort Model is created which computes individual thermal needs with the help of IoT thermal sensor devices. It uses machine learning techniques to predict and provide exact amount of heat and air quality which is needed indoors. An energy management system based on Distributed Energy Resources (DER) uses deep learning algorithms to predict wind speed and solar irradiance. This proposed model helps people to reduce electricity bills, by predicting energy generated by DERs. Figure 16 presents the document citation analysis using WoS for climate sector.

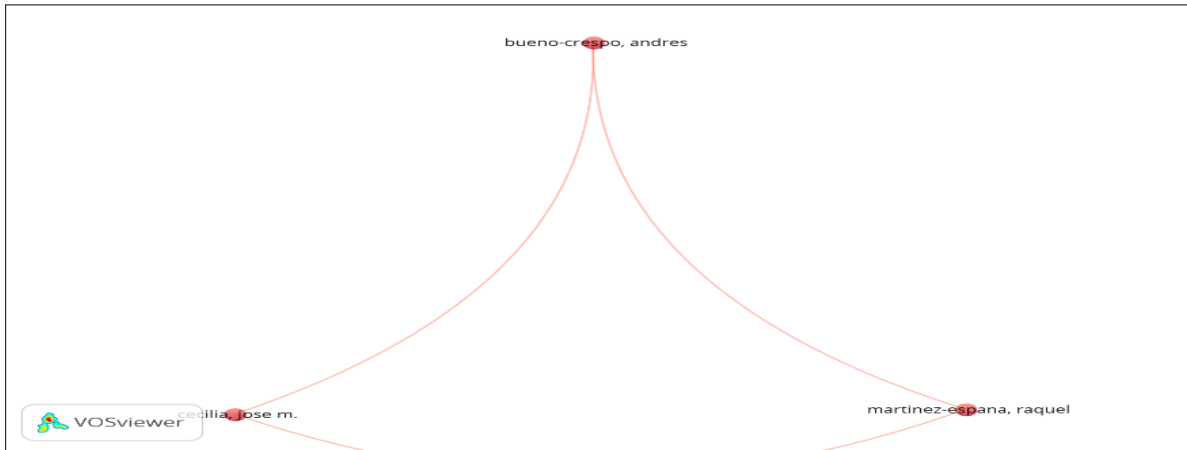


Figure 16: Document citation analysis using WoS data for the Climate sector

To analyze the WoS dataset with respect to countries, the threshold is to be set. The minimum number of documents for a country is set to 5 and the minimum number of citations with respect to a country must be at least 15. Based on these threshold values, 8 out of 44 countries fit the adjusted threshold value. Among which, the United States is at the top with 9 documents and 79 citations. The top 5 countries with their number of published documents and total citations are shown in Table 10.

Table 10. Country-wise citation analysis on the WoS database for Climate Sector

Country	#Documents	#Citations
United States	7	79
England	5	72
India	7	70
Mexico	5	69
Australia	6	64

#### D. Energy Sector

For analysis purpose, strings used for searching publications and the number of outcomes for every search is shown in Table 11.

Table 11: Different datasets used for analysis

S. No.	Data Set	Keyword set for searching	#Papers found
1.	WOS	TOPIC: (Machine Learning) AND TOPIC: (IoT) AND TOPIC: (Energy) Timespan: 2017-2021.	823
2.	Scopus	Title, abstract, keywords: ‘Machine Learning’ AND ‘IoT’ AND ‘Energy’	1629

##### a. Bibliometric Analysis w.r.t Scopus dataset

For bibliometric analysis from the Scopus database, 1629 studies have been extracted. 122 documents are filtered out of 1629, on the basis of most cited (more than 25 citations). Figure 17 represents document citation analysis using Scopus data.

Y. Liu et. al (2019) [24], scored highest in document citation with 313 citations. In this study, deep reinforcement learning technique has been used in energy management system to handle IoT data. Authors have explored various energy management systems that are used in smart cities. Based on these models, an energy management model is proposed in which IoT-based edge computing technology has been used. Machine learning algorithms are used for designing, predicting, and analyzing energy systems. Prediction models have been proposed using machine learning for renewable energy management systems. These models conserve solar energy, wind energy, biofuels etc. Light harvesting systems are being used with the help of solar cells, a sort of IoT device. This device grabs information of surroundings and passes it to artificial neural network (ANN) for inference. ANN helps in prediction using large dataset.

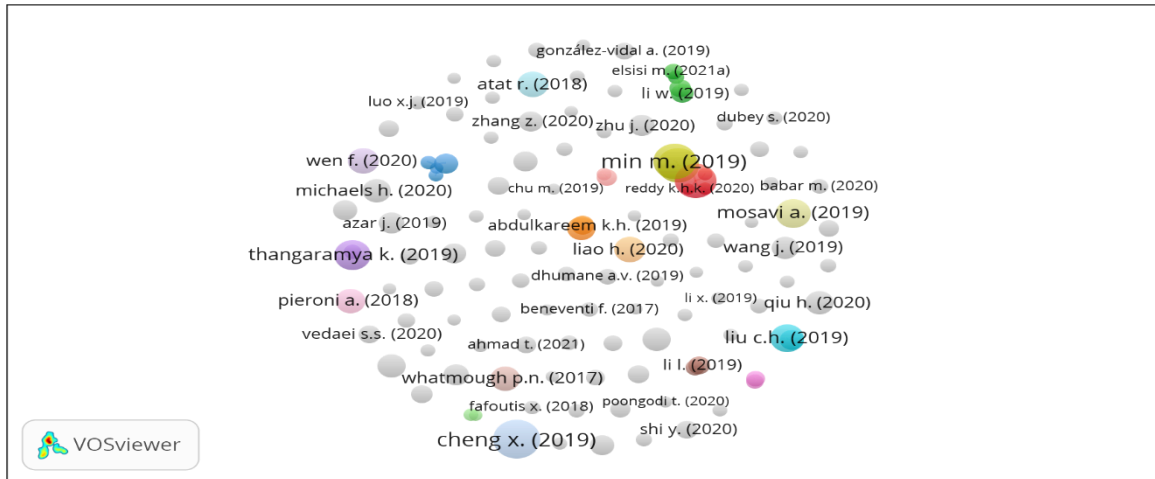


Figure 17: Document citation analysis using Scopus data for the Energy sector

The contribution of different countries in the field of Smart Energy field that uses machine learning and IoT is shown in Figure 18. To get results, the minimum number of documents for a country is set to 5 and the minimum number of citations for a country is set to 15. Using this threshold, 56 out of 126 countries are filtered out, which are further clustered into 11 clusters. Among all countries, the United States is at the top with 3537 citations. The top 5 countries are shown in Table 12, with the number of documents published and total citations.

Table 12. Country-wise citation analysis on the Scopus database for the energy sector

Country	#Documents	#Citations
United States	260	3537
China	196	3322
United Kingdom	100	1817
India	326	1650
Canada	52	1194

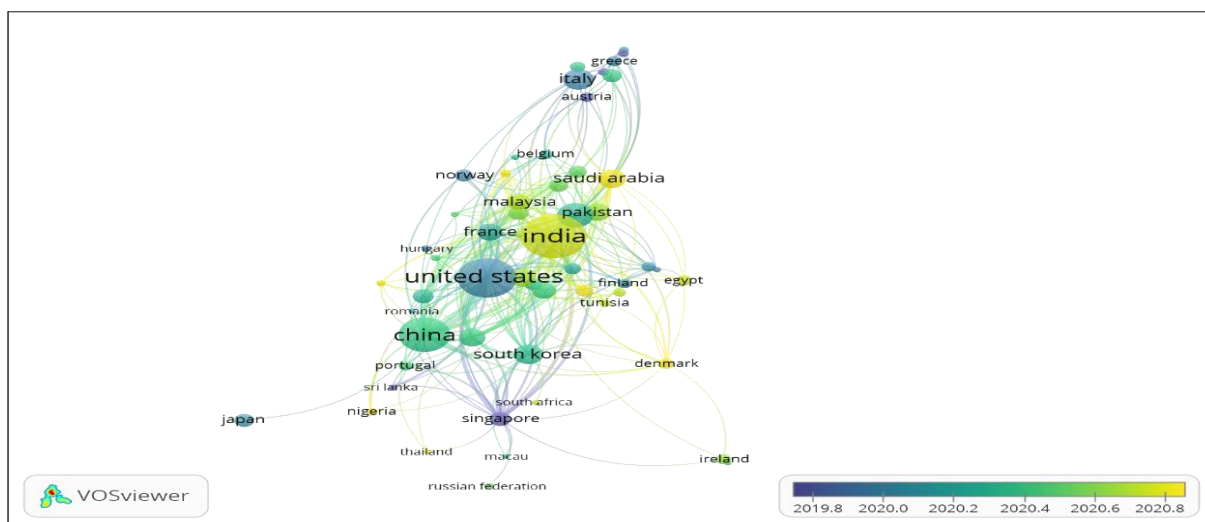


Figure 18: Country-wise citation analysis using Scopus data for the Energy sector

#### b. Bibliometric Analysis w.r.t. Web of Science

For bibliometric analysis from the WoS database, 823 studies have been extracted. 3201 authors contributed a total 823 publications in the research fields of Smart Energy using machine learning. Co-author citation analysis is done by applying a threshold on several documents and some citations. To get the results, the minimum number of documents per author is set to 5 and the minimum number of citations for authors is set to 15. Based on this criterion, the highly cited work of 16 authors is considered for this study. Figure 19 presents the co-author citation analysis results.



### E. Industrial IoT (IIoT) Sector

For analysis purpose, strings used for searching publications and the number of outcomes for every search is shown in Table 14.

Table 14: Different datasets used for analysis

S. No.	Data Set	Keyword set for searching	#Papers found
1.	WOS	TOPIC: (Machine Learning) AND TOPIC: (IoT) AND TOPIC: (Industrial IoT) Timespan: 2017-2021.	606
2.	Scopus	Title, abstract, keywords: ‘Machine Learning’ AND ‘IoT’ AND ‘Industrial IoT’	996

#### a. Bibliometric Analysis w.r.t Scopus dataset

For Bibliometric analysis, only Web of Sciences and Scopus data is used. In Scopus data after the removal of duplicates, 996 publications came as a result of the search. 83 documents out of 996 have been filtered out on the basis of most cited (more than 25 citations) documents, which got clustered on the basis of their connectivity strength. Among these clusters, most cited papers are found and further used to survey different machine learning algorithms applicable to the IIoT sector. Figure 21 represents the document citation analysis.

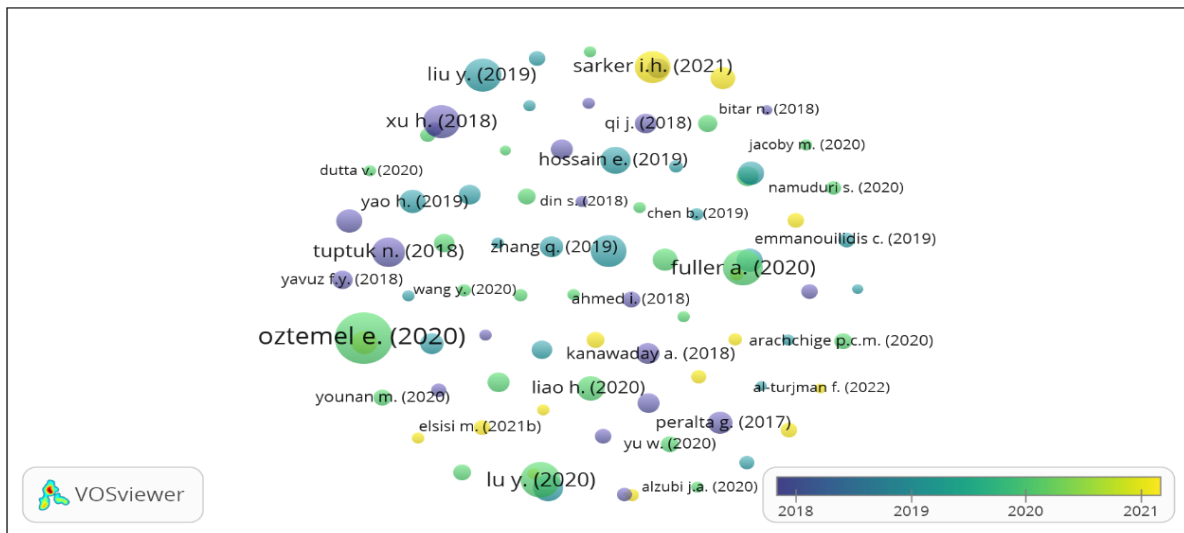


Figure 21: Document citation analysis using Scopus data for the IIoT sector

The word ‘‘Digital twin’’ is introduced, which means a connection between the physical machine and virtual machine through data flow [27]. Industrial 4.0 technology is a workspace where industries use many wearable devices, sensors, and self-decision-making systems using machine learning. Robots dominate most of the manufacturing industries [28]. Most of the cited papers are used to provide a report on the growth of Industry 4.0. A way to secure data for IIoT with integration between machine learning concepts and various present intrusion detection solutions is presented in some studies.

For country-wise citation analysis, a threshold is set as the minimum number of documents of countries should be 5 and the minimum no. of citations for a country is 15. Using this threshold 47 out of 148 countries have been filtered out, which are further clustered into 17 clusters. Among all resulted 47 countries, the United States is at the top with 2670 citations for 137 documents. The top 5 countries with their cited documents are shown in Table 15. Figure 22 shows the most contributed countries in the field of IIoT and used machine learning in their research.

Table 15: Country-wise citation analysis on Scopus database for the Industrial IoT sector

Country	#Documents	#Citations
United States	137	2670
China	44	770
India	90	446
United Kingdom	35	331
Turkey	7	238

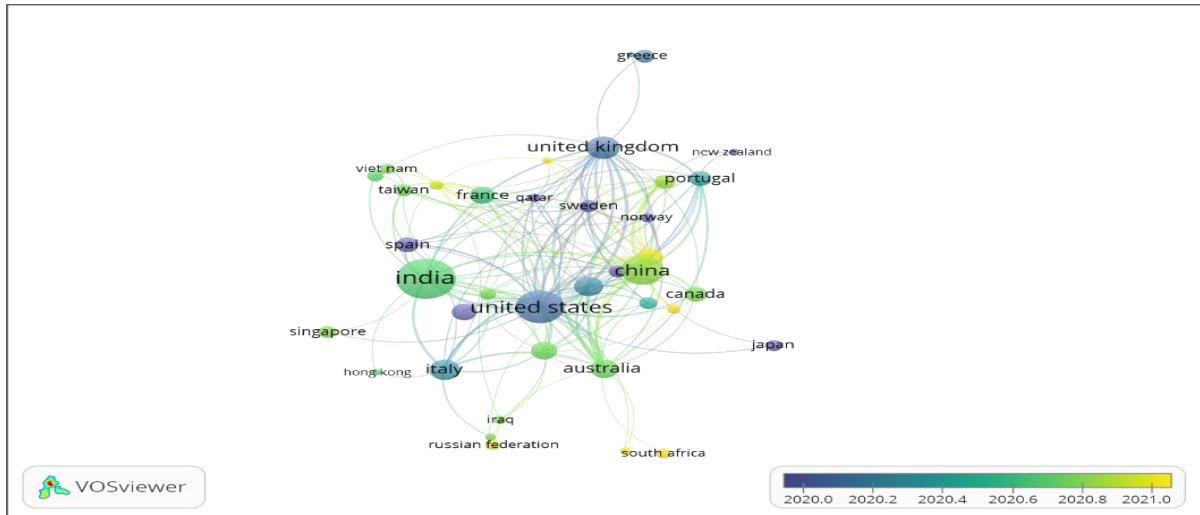


Figure 22: Country-wise citation analysis using Scopus data for the IIoT sector

b. Bibliometric Analysis w.r.t. Web of Science

For bibliometric analysis from the WoS database, 606 studies have been extracted. 2465 authors contributed a total, 606 publications in the research fields of Industrial IoT using machine learning. Co-author citation analysis has been done by applying a threshold on the minimum number of documents per author (set to 5) and a minimum number of citations for authors (set to 15). Based on this criterion, the highly cited work of 11 authors is considered for this study. Figure 23 presents the co-author citation analysis results.

Some of the common indexed studies are already discussed above. A study investigates IIoT areas where big data analytics have made great efforts. Authors have presented the role of machine learning and parallel distributed systems for IIoT data storage and analytics [29]. IIoT Traffic can be classified using correlation for feature selection. To do so, search space must be divided into similar-sized fragments. A hybrid fuzzy C-means clustering algorithm can be used to identify spam over the IIoT cloud network.

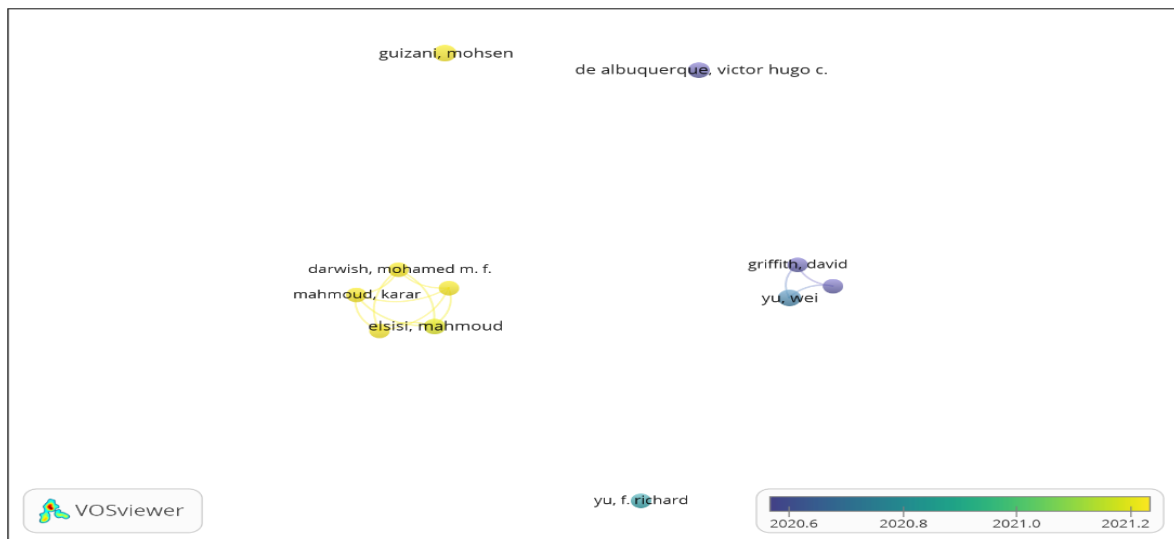


Figure 23: Co-author citation analysis using WoS data for the IIoT sector

Figure 24 shows the most contributed countries in the field of IIoT using machine learning in their research. The threshold is set as the minimum number of documents of countries should be 5 and the minimum no. of citations for a country is 15. Using this threshold, 39 countries have been filtered out from total of 75 countries. These are further clustered into 7 clusters. Among all resulted 39 countries, China is at the top with 2877 citations. The top 5 countries with their total publications and total citations of documents are given in Table 16.

Table 16: Country-wise citation analysis on the WoS database for Industrial IoT sector

Country	#Documents	#Citations
China	160	2877
United States	91	2289
England	53	1155
South Korea	75	1002

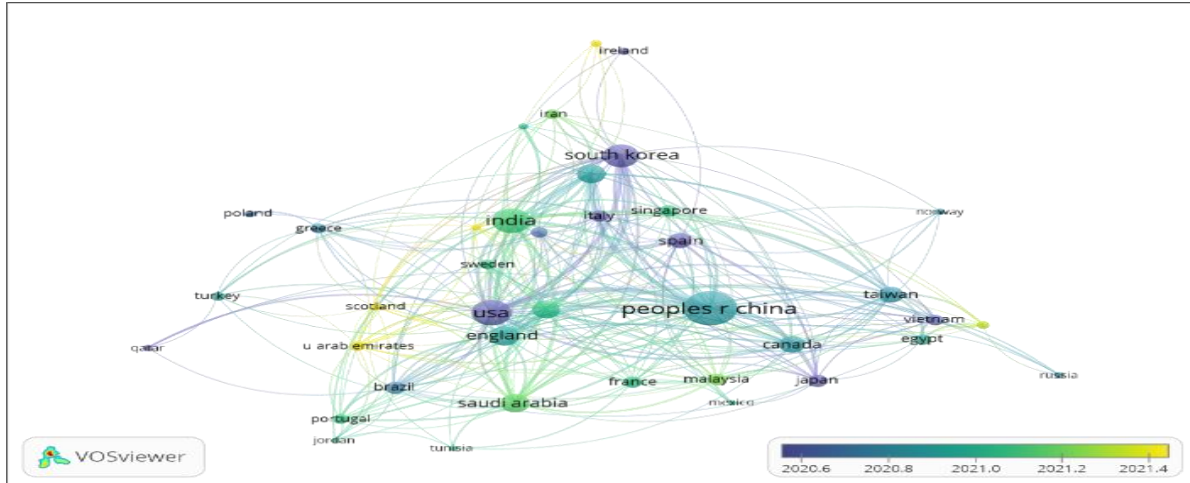


Figure 24: Country-wise citation analysis using WoS data for the IIoT sector

## F. Security Sector

For analysis purpose, strings used for searching publications and the number of outcomes for every search is shown in Table 17.

Table 17: Different datasets used for analysis

S. No.	Data Set	Keyword set for searching	#Papers found
1.	WOS	TOPIC: (Machine Learning) AND TOPIC: (IoT) AND TOPIC: (Security) Timespan: 2017-2021.	1133
2.	Scopus	Title, abstract, keywords: ‘Machine Learning’ AND ‘IoT’ AND ‘Security’	3060

### a. Bibliometric Analysis w.r.t Scopus dataset

For Bibliometric analysis, 3060 articles resulted from the search in the Scopus search engine. Unfortunately, due to the limitation of extracting the data from the Scopus dataset, only 2000 publication details have been downloaded with complete details. 257 documents out of 2000 have been filtered out based on the most cited (more than 25 citations) documents, which have been clustered based on their connectivity strength.

From these resulted publications, most cited papers have been found and further used to survey different machine learning algorithms applicable to the Smart Security sector using IoT. N.C. Luong et. al (2019) [30], publication came out to be the most cited paper in this field with 584 citations. Figure 25 represents the document citation analysis.

Machine learning can help to handle attacks from dynamically changing IoT networks. ML helps to make next-generation IoT a smart and intelligent system. Deep learning algorithms are considered to be much superior to machine learning algorithms in terms of detecting cyber-attacks. Recurrent neural networks (RNN), deep neural networks (DNN), restricted and deep Boltzmann machines, deep belief networks, convolutional neural networks (CNN), and deep autoencoders (DAE) are various deep learning algorithms that are used to classify IoT network data into binary and multiclass [31]. M. H. Fadhil et. al (2019), proposed an Anomaly Detecting IoT system (AD-IoT) to handle network traffic over a smart city IoT network. They used a Random Forest (RF) machine learning algorithm to provide cybersecurity [32].

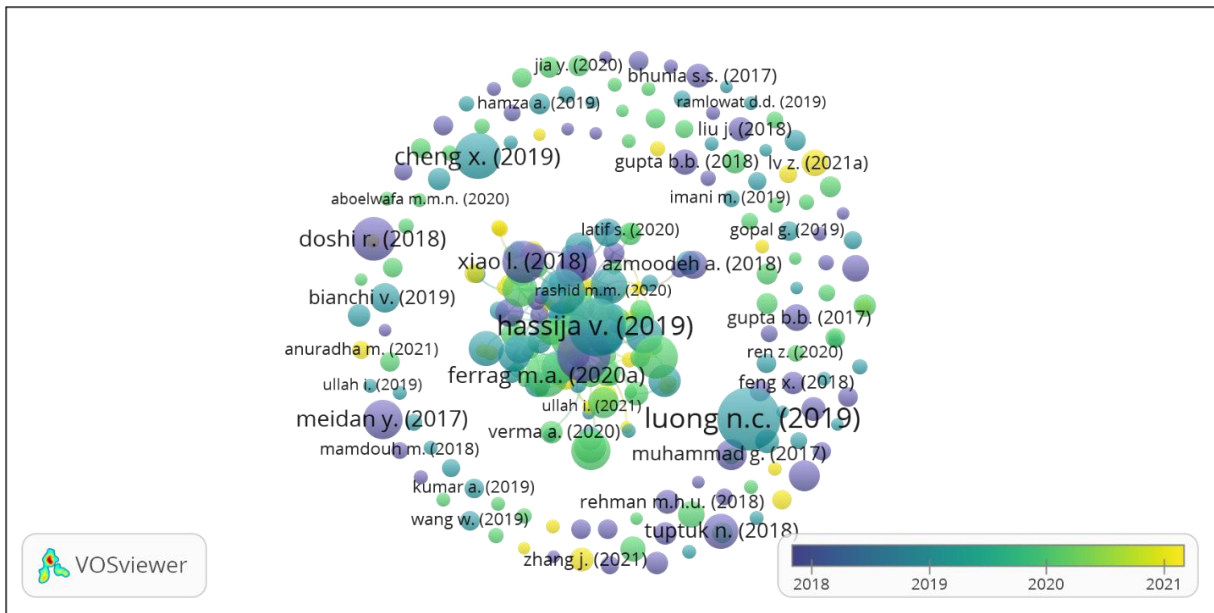


Figure 25: Document citation analysis using Scopus data for the Security sector

For country-wise citation analysis, the threshold is set as minimum documents of countries should be 5 and the minimum no. of citations for a country is 15. Using this threshold 62 out of 157 countries are filtered out, which are further clustered into 8 clusters. Among resulted 62 countries, the United States is at the top with 7474 citations for 386 documents. The top 5 countries' number of publications and citations of documents are shown in Table 18. Figure 26 shows the most contributed countries in the field of Smart Security systems and used machine learning in their research.

Table 18: Country-wise citation analysis on Scopus database for Security Sector

Country	#Documents	#Citations
United States	386	7474
China	266	5980
India	480	4992
Australia	139	3964
United Kingdom	150	2983

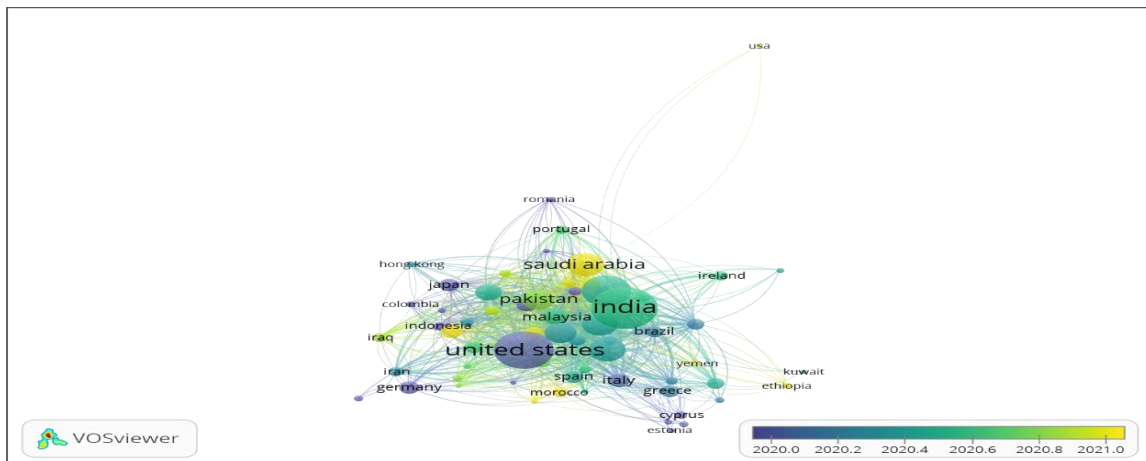


Figure 26: Country-wise citation analysis using Scopus data for the Security sector

b. Bibliometric analysis w.r.t. Web of Science

For bibliometric analysis from the WoS database, 1133 studies have been extracted. 3662 authors contributed a total, of 1133 publications in the research fields of Industrial IoT using machine learning. Co-author citation analysis has been done by applying a threshold on minimum number of documents per author is set to 5 and a minimum number of citations for authors is set to 15. Based on this criterion, the highly cited work of 33 authors is considered for this study. Figure 27 presents the co-author citation analysis results.



## G. Smart City

For analysis purpose, strings used for searching publications and the number of outcomes for every search is shown in Table 20.

Table 20: Different datasets used for analysis

S. No.	Data Set	Keyword set for searching	#Papers found
1.	WOS	TOPIC: (Machine Learning) AND TOPIC: (IoT) AND TOPIC: (Smart City) Timespan: 2017-2021.	378
2.	Scopus	Title, abstract, keywords: ‘Machine Learning’ AND ‘IoT’ AND ‘Smart City’	910

### a. Bibliometric analysis w.r.t Scopus dataset

For Bibliometric analysis, 910 publications came as a result of searching in the Scopus search engine. 88 documents out of 910 have been filtered out on the basis of most cited (more than 25 citations) documents, which are clustered on the basis of their connectivity strength. Among these clusters, the most cited papers have been found and further used to survey different machine learning algorithms applicable to the Smart City sector using IoT and machine learning. Figure 29 represents the document citation analysis. E. Oztemel et. al, 2020 [28] came out to be the most cited document with 576 citations.

B.P.L Lau et. al (2019) [34], surveyed different machine learning techniques that can be used for data fusion for smart city applications. Classification, Regression, and Unsupervised techniques can be used for data fusion. F. Zantalis et. al (2019) [35], reviewed smart transportation for which machine learning and IoT are considered integral parts. Authors include smart lighting, smart parking, accident prevention, and road surveillance as a part of smart transportation systems. The authors discussed ensemble learning, Decision Tree, Clustering, ANN, DL, and linear classification as different machine learning techniques that can be used to manage smart transportation systems [26].

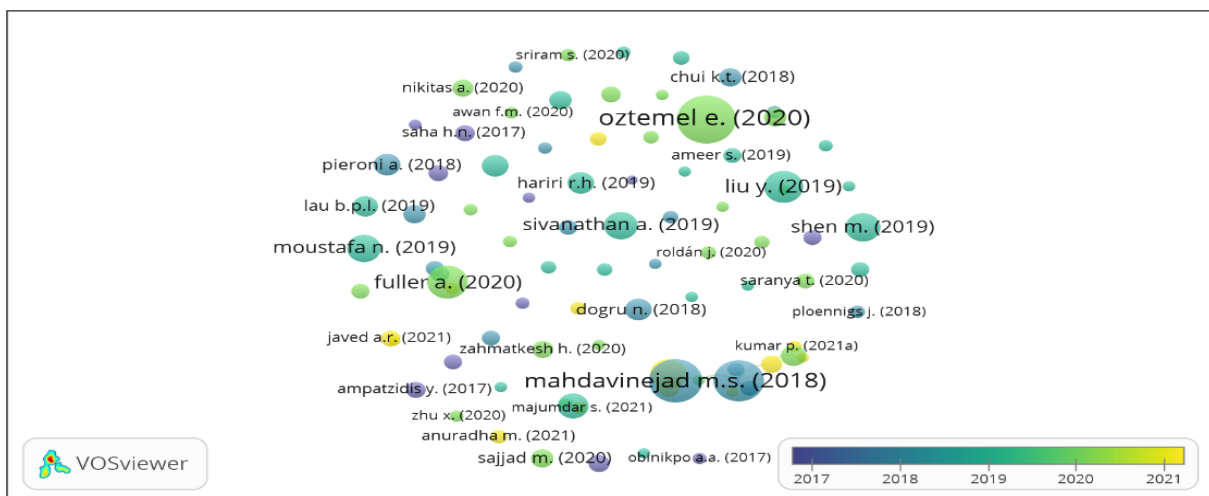


Figure 29: Document citation analysis using Scopus data for the Smart City sector

For country-wise citation analysis, the threshold for minimum documents of countries is set as 5 and the minimum no. of citations for a country is set as 15. Using this threshold, 38 out of 101 countries have been filtered out, which are further clustered into 8 clusters. Among all resulted 38 countries, the United States is at the top with 3107 citations for 119 documents. The top 5 countries with total publications in this field and total citations are shown in Table 21. Figure 30 shows the most contributed countries in the field of Smart Cities that have used machine learning in their research.

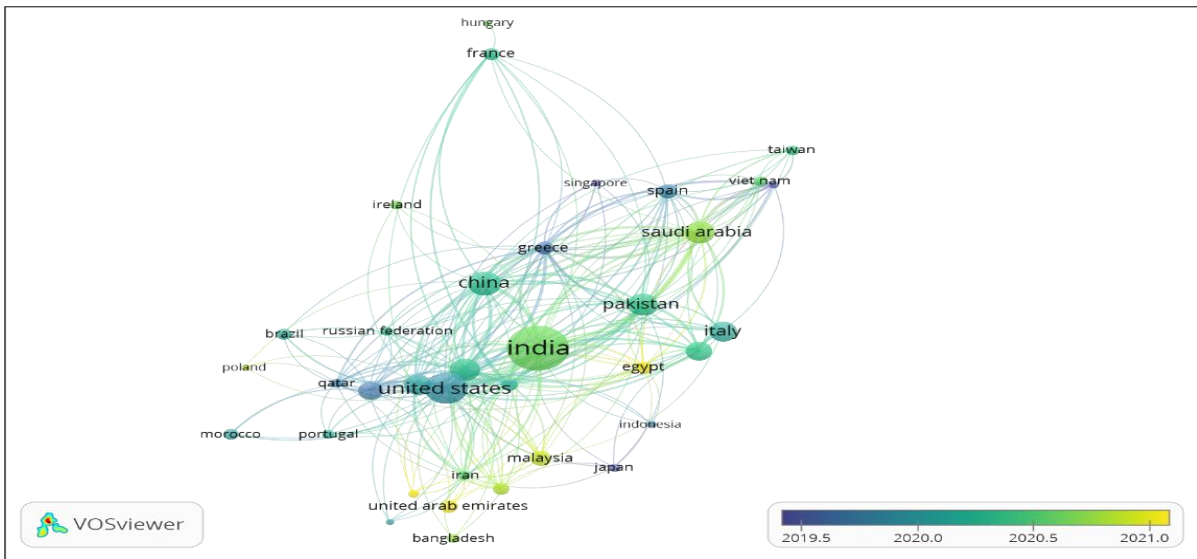


Figure 30: Country-wise citation analysis using Scopus data for the Smart City sector

Table 21: Country-wise citation analysis on Scopus database for Smart City

Country	#Documents	#Citations
United States	119	3107
Australia	47	1625
China	69	1564
United Kingdom	59	1558
India	240	1043

b. Bibliometric analysis w.r.t. Web of Science

For bibliometric analysis from the WoS database, 378 studies have been extracted. 1564 authors contributed a total, 378 publications in the research fields of Smart City using machine learning. Co-author citation analysis is done by applying a threshold on, a minimum number of documents per author is set to 5 and a minimum number of citations for authors is set to 15. Based on this criterion, the highly cited work of 4 authors is considered for this study. Figure 31 presents the co-author citation analysis results.



Figure 31: Co-author citation analysis using WoS data for the Smart City sector

Five studies are indexed in both Scopus and WoS databases. Data fusion techniques for smart cities, autonomous vehicles, etc. can be improved with various mathematical methods like artificial intelligence, probabilistic methods, etc. The authors proposed a hybrid model using ARIMA and Bi-LSTM optimized using Aquila

optimizer. This proposed hybrid model is used to handle traffic congestion problems. for Different machine learning algorithms can be used for mining purposes over IoT datasets. Among them, Deep Learning ANNs, C4.5, and C5.0 gave the best accuracy results. Every sector of smart cities like smart health management systems, smart traffic management systems, unmanned aerial vehicles, etc. is likely to evolve with Artificial Intelligence (AI), Deep Reinforcement Learning (DRL), and Machine Learning (ML) technologies only [36].

Figure 32 shows the most contributed countries in the field of IoT Smart City and used machine learning in their research. The threshold is set as the minimum number of documents of countries should be 5 and the minimum no. of citations for a country is 15. Using this threshold 33 countries are filtered out from total 74 countries. These are further clustered into 6 clusters. Among all resulted 33 countries, the USA is at the top with 1786 citations. The top 5 countries with their cited documents are given in Table 22.

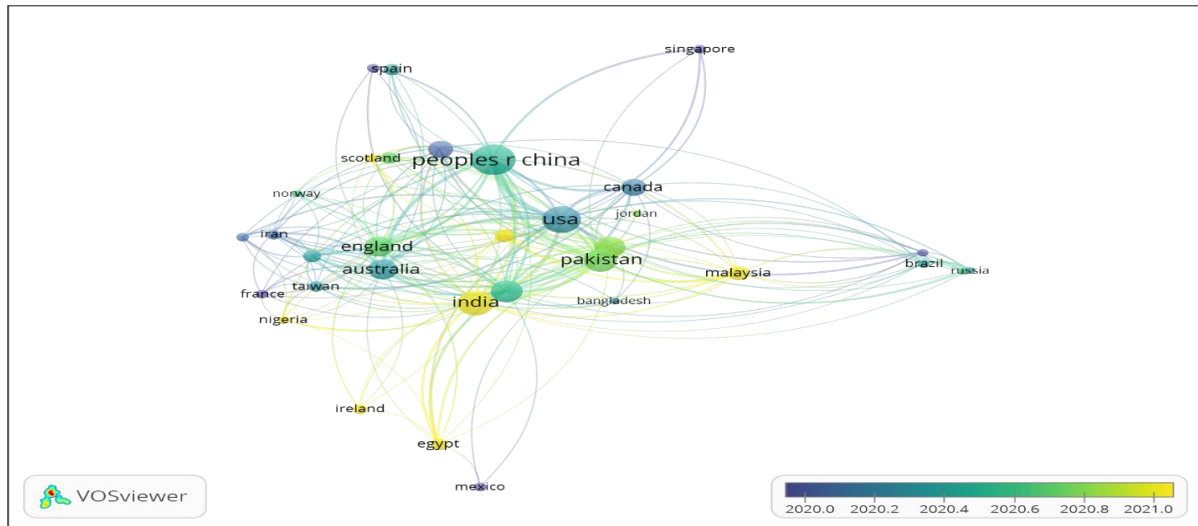


Figure 32: Country-wise citation analysis using WoS data for the IoT Smart City sector

Table 22: Country-wise citation analysis on WoS database for Smart City

Country	#Documents	#Citations
United States	65	1786
China	84	1541
Australia	37	1271
England	39	1257
Saudi Arabia	47	902

#### 4. Result Outcomes

This section discusses results based on bibliometric analysis and the limitations of the study. The sector-wise bibliometric analysis will help in finding the kind of machine learning algorithms used worldwide for a particular IoT application sector. This study finds the majority of quality publications that can be helpful for researchers before moving in any direction of their research work. Table 23 and Table 24 summarize the results of the highly cited papers and countries respectively, for each application domain of IoT where machine learning plays an integral part of the research.

Table 23: Most cited paper for different IoT domains using ML techniques.

S. No.	IoT Domain	Publication	#Citation
1.	Healthcare	R. Vaishya <i>et. al</i> , 2020 [9]	523
2.	Agriculture	K. Jha <i>et. al</i> , 2019 [17]	243
3.	Climate	J. Pang <i>et. al</i> , 2021 [20]	43
4.	Energy	Y. Liu <i>et. al</i> , 2019 [24]	313
5.	IIoT	E. Oztemel <i>et. al</i> , 2020 [28]	576
6.	Security	N.C. Luong <i>et. al</i> , 2019 [30]	584
7.	Smart City	E. Oztemel <i>et. al</i> , 2020 [28]	576

Table 24: most contributed countries in different IoT domains using ML techniques

S. No.	IoT Domain	Most contributed country	#Citation
--------	------------	--------------------------	-----------

1.	Healthcare	India	3916
2.	Agriculture	India	2090
3.	Climate	India	169
4.	Energy	United States	3537
5.	IIoT	China	2877
6.	Security	United States	7474
7.	Smart City	United States	3107

After going through all the results, it is seen that IoT domains like Healthcare, Energy, Security, and Smart City have a lot of academic literature in the databases as compared to Climate, agriculture, and IIoT. India is a country having maximum citations for three IoT sectors i.e. healthcare, agriculture, and Climate. The United States is also a leading country with the highest citation values under three IoT domains i.e. Energy, Security, and Smart City. China has the most cited research work for the IIoT application area of IoT and machine learning.

After reviewing filtered articles, this study can find the kind of data used in every IoT application sector. Also, it provides a kind of machine learning algorithm that can be used to manage the data for its respective IoT application sector. The outcomes of the study are summarised in Table 25. It is not possible to review such a large number of publications and produce a textual analysis for it. So, systematic literature review techniques is used to extract some important articles for analysis. Papers from the WoS and Scopus are reviewed thoroughly after filtering them from thousands of articles that are extracted from the database. For this review, data for only the past 6 years are extracted still, it gave results in thousands for one IoT application sector. Some of these papers got repeated for more than one sector, but these repeated papers have been kept to one sector only, for which they are found to be more relevant. This process is done manually.

Table 25: Summarized Outcomes

S.No.	IoT application sector	Kind of data used	Machine Learning Algorithms/ technique	Referenced articles
1.	Healthcare	Real-Time Data, Temporal data, Genomic data, High dimensional data	SVM, Naive Bayes, Random Forest, Decision Tree, KNN	[13; 14; 15]
2.	Agriculture	Temporal data, real-time data	KNN, deep reinforcement learning, linear regression, Bayesian network	[18; 19]
3.	Climate	Real-Time Data, Spatiotemporal data	Deep learning, supervised machine learning	[22; 23; 37]
4.	Energy	High dimensional data, temporal data	ANN, CNN, KNN, reinforcement learning	[25; 26; 40]
5.	Industrial IoT	Real time data	reinforcement learning, KNN, SVM, ANN, naïve Bayes, Logistic regression	[28; 29; 39]
6.	Security	Real-time data	decision tree, NB, ANN, RF, DNN, RNN, DAE, CNN, deep belief network	[32; 38]
7.	Smart City	Real-Time Data, Temporal data	SVM, regression, ensemble learning, Decision tree, Clustering, ANN, DL, RF	[25; 35; 36]

The limitation of this study is that it is based on citation analysis only. Though Citation may help to pick the most popular article in the field but bypasses the recent papers as recent papers may not have high citation value. Data collected for this study is limited to the searched keywords only. Also, only two databases study i.e. Scopus and WoS are used for this study. Other databases like IEEE Xplore, Springer, etc. can also be used in the future for this kind of bibliometric review.

## 5. Conclusion

Machine learning has evolved so rapidly and started influencing everyday life of individuals apart from all segments contributing in realizing Smart city projects across the globe. To analyze such a large amount of data, machine learning techniques and IoT makes a great duo. This paper elaborates on the connection between machine learning and IoT concerning different IoT application areas. In this study, Scopus and WoS databases are used to extract published articles related to machine learning and IoT. To handle such a large dataset, Bibliometric analysis is done. Systematic Literature Review steps have been followed to perform bibliometric analysis on the extracted articles. These articles are then filtered based on author and country-wise citations. These filtered articles are then reviewed thoroughly to get in-depth knowledge of IoT and ML technology conflux. Some of the articles discussed in this study give the impression that deep learning algorithms outperform in some fields as compared to machine learning algorithms. This study concluded the kind of data one has to deal with, for a particular IoT application sector. It also elaborates on multiple machine learning techniques that are suitable to manage data for particular application sectors of IoT. This study helps to conclude the majority kind of data types that one needs to deal with for a particular IoT application domain. This study will allow researchers and practitioners to generate a well-defined road map to use machine learning techniques for particular IoT application areas.

**Funding:** “This research received no external funding”

**Conflicts of Interest:** “The authors declare no conflict of interest.”

## References

- [1] Panthi, V., Mishra, A. K., 2023. Enhancing Healthcare Monitoring through the Integration of IoT Networks and Machine Learning. *Journal of International Journal of Wireless and Ad Hoc Communication*, 7(1), 28-39. <https://doi.org/10.54216/IJWAC.070103>
- [2] Injadat, M., Moubayed, A., Nassif, A.B., Shami, A., 2021. Machine learning towards intelligent systems: applications, challenges, and opportunities. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-020-09948-w>.
- [3] Zyoud, S.H., Al-Jabi, S.W., Amer, R., Shakhshir, M., Shahwan, M., Jairoun, A.A., Akkawi, M., Abu Taha, A., 2022. Global research trends on the links between the gut microbiome and cancer: a visualization analysis. *Journal of Translational Medicine* 20, 83. <https://doi.org/10.1186/s12967-022-03293-y>
- [4] Ligthart, A., Catal, C., Tekinerdogan, B., 2021. Systematic reviews in sentiment analysis: a tertiary study. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-021-09973>
- [5] Casillas, J., Acedo, F., 2007. Evolution of the Intellectual Structure of Family Business Literature: A Bibliometric Study of FBR. *Family Business Review* 20, 141–162. <https://doi.org/10.1111/j.1741-6248.2007.00092.x>
- [6] Talboom, J.S., Huentelman, M.J., 2018. Big data collision: the internet of things, wearable devices and genomics in the study of neurological traits and disease. *Human Molecular Genetics* 27, R35–R39. <https://doi.org/10.1093/hmg/ddy092>
- [7] Pallepati Vasavi, A. Punitha, T. Venkat Narayana Rao, 2023. Chili Leaf Disease Detection Using Deep Feature Extraction, *Journal of Intelligent Systems and Internet of Things*, 9(2), 222-230. <https://doi.org/10.54216/JISIoT.090216>
- [8] Myvizhi M., Ahmed Abdel-Monem, 2023. Wind Turbine Prediction using Deep Learning and Long Short Term Memory (LSTM), *International Journal of Advances in Applied Computational Intelligence*, 3(2), 48-57. <https://doi.org/10.54216/IJAACI.030205>.
- [9] Vaishya, R., Javaid, M., Khan, I.H., Haleem, A., 2020. Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 14, 337–339. <https://doi.org/10.1016/j.dsx.2020.04.012>
- [10] Neu, D.A., Lahann, J., Fettke, P., 2021. A systematic literature review on state-of-the-art deep learning methods for process prediction. *Artificial Intelligence Review*, 55, 801-827. <https://doi.org/10.1007/s10462-021-09960-8>
- [11] Li, J.-P.O., Liu, H., Ting, D.S.J., Jeon, S., Chan, R.V.P., Kim, J.E., Sim, D.A., Thomas, P.B.M., Lin, H., Chen, Y., Sakamoto, T., Loewenstein, A., Lam, D.S.C., Pasquale, L.R., Wong, T.Y., Lam, L.A., Ting, D.S.W., 2021. Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective. *Progress in Retinal and Eye Research* 82, 100900. <https://doi.org/10.1016/j.preteyeres.2020.100900>

- [12] Javed, A.R., Fahad, L.G., Farhan, A.A., Abbas, S., Srivastava, G., Parizi, R.M., Khan, M.S., 2021. Automated cognitive health assessment in smart homes using machine learning. *Sustainable Cities and Society* 65, 102572. <https://doi.org/10.1016/j.scs.2020.102572>
- [13] David, S., Andrew, J., Sagayam, K. M., Elngar, A. A., 2021. Augmenting security for electronic patient health record (ePHR) monitoring system using cryptographic key management schemes. *Journal of Fusion: Practice and Applications*, 5 (2), 51-61. <https://doi.org/10.54216/FPA.050201>
- [14] Chui, K.T., Alhalabi, W., Pang, S.S.H., Pablos, P.O. de, Liu, R.W., Zhao, M., 2017. Disease Diagnosis in Smart Healthcare: Innovation, Technologies and Applications. *Sustainability* 9, 2309. <https://doi.org/10.3390/su9122309>
- [15] Kaur, P., Kumar, R., Kumar, M., 2019. A healthcare monitoring system using random forest and Internet of Things (IoT). *Multimedia Tools Application*, 78, 19905–19916. <https://doi.org/10.1007/s11042-019-7327-8>
- [16] Jain, A., Gupta, J., Khandelwal, S., Kaur, S., 2021. Vehicle License Plate Recognition. *Journal of Fusion: Practice and Applications*, 4 (1), 15-21. <https://doi.org/10.54216/FPA.040102>
- [17] Jha, K., Doshi, A., Patel, P., Shah, M., 2019. A comprehensive review on automation in agriculture using artificial intelligence. *Artificial Intelligence in Agriculture* 2, 1–12. <https://doi.org/10.1016/j.aiia.2019.05.004>
- [18] Balducci, F., Impedovo, D., Pirlo, G., 2018. Machine Learning Applications on Agricultural Datasets for Smart Farm Enhancement. *Machines* 6, 38. <https://doi.org/10.3390/machines6030038>
- [19] Taştan, M., Gökozan, H., 2019. Real-Time Monitoring of Indoor Air Quality with Internet of Things-Based E-Nose. *Applied Sciences* 9, 3435. <https://doi.org/10.3390/app9163435>
- [20] Pang, J., Huang, Y., Xie, Z., Han, Q., Cai, Z., 2021. Realizing the Heterogeneity: A Self-Organized Federated Learning Framework for IoT. *IEEE Internet of Things Journal* 8, 3088–3098. <https://doi.org/10.1109/JIOT.2020.3007662>
- [21] Ponce, H., Gutiérrez, S., 2019. An indoor predicting climate conditions approach using Internet-of-Things and artificial hydrocarbon networks. *Measurement* 135, 170–179. <https://doi.org/10.1016/j.measurement.2018.11.043>
- [22] Aguilera, J.J., Kazanci, O.B., Toftum, J., 2019. Thermal adaptation in occupant-driven HVAC control. *Journal of Building Engineering* 25, 100846. <https://doi.org/10.1016/j.jobe.2019.100846>
- [23] Alhussein, M., Haider, S.I., Aurangzeb, K., 2019. Microgrid-Level Energy Management Approach Based on Short-Term Forecasting of Wind Speed and Solar Irradiance. *Energies* 12, 1487. <https://doi.org/10.3390/en12081487>
- [24] Liu, Y., Yang, C., Jiang, L., Xie, S., Zhang, Y., 2019. Intelligent Edge Computing for IoT-Based Energy Management in Smart Cities. *IEEE Network* 33, 111–117. <https://doi.org/10.1109/MNET.2019.1800254>
- [25] El-Wakeel, A.S., Li, J., Noureldin, A., Hassanein, H.S., Zorba, N., 2018. Towards a Practical Crowdsensing System for Road Surface Conditions Monitoring. *IEEE Internet of Things Journal* 5, 4672–4685. <https://doi.org/10.1109/JIOT.2018.2807408>
- [26] Manikandan, R., Keerthana, S., Priya, S. S., Madhumitha, R., Aditya, A. G. S., Priya, D., 2023. Android-based System for Intelligent Traffic Signal Control and Emergency Call Functionality, *Journal of Journal of Cognitive Human-Computer Interaction*, 5(2), 31-44. <https://doi.org/10.54216/JCHCI.050204>
- [27] Pustokhin, D. A., Pustokhina, I. V., 2023. FLC-NET: Federated Lightweight Network for Early Discovery of Malware in Resource-constrained IoT, *International Journal of Wireless and Ad Hoc Communication*, 6(2)43-55. <https://doi.org/10.54216/IJWAC.060204>
- [28] Oztemel, E., Gursev, S., 2020. Literature review of Industry 4.0 and related technologies. *Journal of Intelligent Manufacturing*, 31, 127–182. <https://doi.org/10.1007/s10845-018-1433-8>
- [29] Rehman, M.H. ur, Ahmed, E., Yaqoob, I., Hashem, I.A.T., Imran, M., Ahmad, S., 2018. Big Data Analytics in Industrial IoT Using a Concentric Computing Model. *IEEE Communications Magazine* 56, 37–43. <https://doi.org/10.1109/MCOM.2018.1700632>
- [30] Luong, N.C., Hoang, D.T., Gong, S., Niyato, D., Wang, P., Liang, Y.-C., Kim, D.I., 2019. Applications of Deep Reinforcement Learning in Communications and Networking: A Survey. *IEEE Communications Surveys & Tutorials* 21, 3133–3174. <https://doi.org/10.1109/COMST.2019.2916583>

- [31] Ferrag, M.A., Maglaras, L., Moschoyiannis, S., Janicke, H., 2020. Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications* 50, 102419. <https://doi.org/10.1016/j.jisa.2019.102419>
- [32] Fadhil, M. H., Makhool, N.Q., Hummady, M. M., Dawood, Z. O., 2022. Machine Learning-based Information Security Model for Botnet Detection, *Journal of Cybersecurity and Information Management*, 9(1), 68-79, <https://doi.org/10.54216/JCIM.090106>
- [33] Thamilarasu, G., Chawla, S., 2019. Towards Deep-Learning-Driven Intrusion Detection for the Internet of Things. *Sensors* 19, 1977. <https://doi.org/10.3390/s19091977>
- [34] Lau, B.P.L., Marakkalage, S.H., Zhou, Y., Hassan, N.U., Yuen, C., Zhang, M., Tan, U.-X., 2019. A survey of data fusion in smart city applications. *Information Fusion* 52, 357–374. <https://doi.org/10.1016/j.inffus.2019.05.004>
- [35] Zantalis, F., Koulouras, G., Karabetsos, S., Kandris, D., 2019. A Review of Machine Learning and IoT in Smart Transportation. *Future Internet* 11, 94. <https://doi.org/10.3390/fi11040094>
- [36] Ullah, Z., Al-Turjman, F., Mostarda, L., Gagliardi, R., 2020. Applications of Artificial Intelligence and Machine learning in smart cities. *Computer Communications* 154, 313–323. <https://doi.org/10.1016/j.comcom.2020.02.069>
- [37] Sanjay Rajan J., Sanjay Kumar M., Madesh G., Sanjai S., C. Manikandan, 2023. Online Weather Station, *Journal of Journal of Cognitive Human-Computer Interaction*, Vol. 5(2), 24-30. <https://doi.org/10.54216/JCHCI.050203>
- [38] Lobna Osman, 2022. An Intelligent Spatial Military Intrusion Detection using Reactive Mobility Unmanned Vehicles Based on IoT and metaheuristic Optimization Algorithm, *Journal of Cybersecurity and Information Management*, 9(2), 31-41. <https://doi.org/10.54216/JCIM.090203>
- [39] Ossama H. Embarak, Raed Abu Zitar, 2023. Securing Wireless Sensor Networks Against DoS attacks in Industrial 4.0, *Journal of Intelligent Systems and Internet of Things*, 8(1), 66-74. <https://doi.org/10.54216/JISIoT.080106>
- [40] Tamer H. M. Soliman, 2023. Neutrosophic Multi-Criteria Decision Making COMET Method for Evaluation Sustainable Electricity Generation Considering Renewable Energy Sources, *International Journal of Advances in Applied Computational Intelligence*, 4 (1), 19-27. <https://doi.org/10.54216/IJAACI.040102>