Group_18_Project_report.pdf

by

Bibliometrics & Citation Analysis of Pakistani Researchers in the Field of Computer Science

ALI KHALID[1], MUHAMMAD AHMAD[2], MUHAMMAD TAYYAB[3], AND HAIDER ALI ADEEL[4]

[1]Information Technology University, Lahore, Pakistan (e-mail: masds21001@itu.edu.pk)
[2]Information Technology University, Lahore, Pakistan (e-mail: masds21021@itu.edu.pk)
[3]Information Technology University, Lahore, Pakistan (e-mail: masds21032@itu.edu.pk)
[4]Information Technology University, Lahore, Pakistan (e-mail: masds21044@itu.edu.pk)

ABSTRACT The study of assessing scientific research patterns and trends in a certain community, country, or area is known as bibliometric analysis. The bibliometric analysis of all computer science-related publications by Pakistani authors is carried out in this work. The information is gathered from a variety of sources on the web of science. Indexes. Annual Scientific Production, Average Citation, and other factors are examined. as stated in the article The Most Cited Documents, the Most Relevant Authors, the Most Relevant Sources (Journals), and the Most Relevant Publishers popular subjects, for example. For a better understanding, a content-based, citation-based, and time series-based analysis is carried out. grasp of current publishing tendencies The pace of research has been modest in recent years, but it has picked up in recent years2014-2020,It grows at an exponential rate. With over 1400 publications, IEEE Access is the most relevant journal. China, Korea, Malaysia, and Saudi Arabia have the most corresponding authors. The major research funding agencies are China's National Natural Science Foundation, the Higher Education Commission, and King Saud University. The most popular author keywords are Machine Learning, Deep Learning, and Cloud Computing, which are all hot subjects in recent years. Articles make up the majority of the papers, followed by Reviews and Articles; Early Access. Before 2004, the average article citation was quite low; after 2004, it increased by 50%.In contrast, more than 38% of research publications receive fewer than 5 citations. Only 3% of the publications received more than 50 citations, and 28% of the publications received no citations.

INDEX TERMS Bibliometrics, Citation analysis, Research Impact, Computer Science, Pakistan

I. INTRODUCTION

The scene of science and exploration is quickly developing. The number of people joining the field of research and innovation is increasing with the increase in population. Gone are the days when all individuals from a college office would praise the fruitful distribution of a colleague's paper [1]. Earlier, researchers would just consider the number of papers they had distributed as a proportion of their scholastic standing. The evolution of science has also led to the evolution of the quality of research work. Today the center is progressively moving from whether an analyst has published a paper to where he/she has published it and the effect that piece of exploration has on established researchers and the world at large [2].

What criteria do you use to evaluate the quality of a research paper? How can you tell if your research is having an impact and is considered significant? An objective way is through citation analysis. Citation analysis is the process of evaluating and interpreting the citations that articles, scientists, universities, and countries receive to determine their scientific influence and productivity. Any research article should include a list of references that directs readers to previous relevant research. A reference, often known as a citation, is a type of recognition given by one research work to another. Scientists build on previous work to uncover new knowledge. Researchers read relevant published research and use it as a foundation for arguments made in their research papers to find gaps in existing research and choose a research topic.

Today's researchers are under increased pressure to publish their findings. Academic departments are required to publish a certain number of articles each year. Both individ-

uals and institutions have a lot riding on the evaluation of research quality. As a result, governments, funding agencies, and tenure and promotion committees are looking for simple and objective approaches to evaluate growing research volumes in the shortest amount of time possible. To this purpose, they are increasingly resorting to citation analysis for objective impact assessment metrics. Data science is a hot research topic in this decade. Therefore, in this work, we performed a citation analysis of the publications in the field of data science by Pakistani authors. This analysis will help the researchers to identify trending topics, famous journals and productive research groups. Moreover, they will be in a better position to understand the reach of their work, identify the patterns in which their research work is used, benchmark themselves against their peers, and set objectives for themselves and their publications.

II. LITERATURE REVIEW

A citation index is similar to a bibliographic index in that it is a list of citations between publications that allows the user to quickly determine which later papers cite which older documents. There are different citation index databases. Some of them are proprietary while others are freely accessible to the public. The scope of these databases also varies. Most of them are focused on a single field of research and index material particular to that field. PubMed [6](Medical and biomedicine research), Abstracts of Chemicals [7], Mathematical Testimonials [8], the ACM Library on the Internet [9] (computer sciences), and CiteSeer [10] (computer and information sciences), are some of the specialist databases that are accessible over the internet and allow open access to bibliometric data in their field. Contrary to this, Web of Science [11](Clarivate Analytics) and Scopus [12] (Elsevier) are multidisciplinary commercial citation index databases and most of the bibliometric studies are based on data covering these two databases. Google Scholar [15] is another free web search engine that indexes the full text or metadata of scholarly literature published in a variety of formats and fields. Researchers have also performed studies on the comparative analysis and coverage of databases [13], [14].

Citation impact uses citation indexes to find the number of times an academic journal article, book, or author is cited by other articles, books, or writers [3]–[5]. Citation counts are used to determine the impact or influence of academic work, giving rise to the area of bibliometrics or scientometrics, which focuses on the study of academic impact patterns through citation analysis. Different citation index databases has introduced different metrics to measure citation impacts. The journal impact factor is a measure of the relevance of journals based on the two-year average ratio of citations to papers published. Academic institutions use it to make judgments about academic tenure, promotion, and employment while authors use it to choose which journal to publish in. The impact factor (IF) or journal impact factor (JIF) of an academic journal measured by Clarivate's Web of Science is a scientometric indicator that quantifies the yearly mean number of citations of papers published in the previous two years in a certain magazine. It is widely employed as a proxy for a journal's relative relevance within its area as a journal-level indicator; journals with higher impact factor values are accorded the status of being more important, or carrying more prestige in their respective disciplines, than journals with lower values. While it is widely used by universities and funding organizations to make decisions about promotion and research proposals, it has recently been accused of distorting sound scientific standards [16], [17]. Journal rank is another metric which is extensively used in academic circles to assess the influence and quality of academic journals. The purpose of journal rankings is to represent a journal's position within its area, the relative difficulty of publishing in that journal, and the prestige associated with it. In a number of countries, they have been adopted as official research evaluation instruments [18]. CiteScore (CS) [19] is an academic publication's metric that reflects the yearly average number of citations to recent articles published in the journal. Elsevier introduced this journal rating statistic in December 2016 as an alternative to the commonly used JCR impact factors (calculated by Clarivate). CiteScore is based on citations recorded in the Scopus database rather than the JCR database, and those citations are collected for publications published in the previous four years rather than the previous two or five years. The SCImago Journal Rank (SJR) indicator is a measure of academic journals' scientific influence that takes into consideration both the number of citations received and the prominence or prestige of the journals from which the citations are obtained. The SJR indicator [20] of a journal is a numeric value that represents the average number of weighted citations received per document published in that journal during the previous three years, as indexed by Scopus, during a given year. The higher the SJR indicator value, the more prestigious the journal. The Scimago Lab, which began as a research group at the University of Granada, produced SJR.

Iqbal et al. [21] examine the computer networking domain and present a study based on the content and metadata of four major computer networking journals; IEEE Communications Surveys and Tutorials (COMST), IEEE/ACM Transactions on Networking (TON), ACM Special Interest Group on Data Communications (SIGCOMM), and IEEE International Conference on Computer Communications (INFOCOM). They compared the publication trends in INFOCOM and SIG-COMM to the co-evolution of trends in the COMST and TON journals. They used metadata analysis, content-based analysis, and citation analysis to analyze the computer networking literature. They also identified noteworthy trends as well as the most influential authors, institutes, and nations based on the number of publications and article citations.

Through a survey, Ali et al. [22] conducted a bibliometric analysis of Pakistani Library and Information Sciences (LIS) scholars' research. The authors created a questionnaire which they circulated to LIS scholars via email, Yahoo groups, and Facebook to representatives from all over

Pakistan. The data from a total of 104 respondents were then analyzed using the SPSS version 21 programme. They identified relevant demographic information; gender and location, the extent of collaborative authorship, the extent of publishing based on geographical regions, the strength of the association between job title(seniority) and the number of publications, the strength of citation metrics for national outputs, and factors that may harm LIS scholars' ability to conduct research and/or publish it.

For the period 1980–2011, R. S. Bajwa et al. [23] examines the research trends in Pakistan in the field of biotechnology. They looked at the rise and fall of the annual growth rate, as well as the comparison of organizations that actively participate in biotechnology research using the publication rate, citation rate, average citation per paper, and numerous indexing methods. The top 100 articles in the discipline of software engineering were computed and classified by Garousi et al. [24]. Based on the the total number of citations and the average number of citations per year, they discovered the best publications. They established a GQM (Goal, Question, and Metric) system to determine their goal by formulating certain questions. They also compared the top papers to the top papers from other fields of study. They also determined which fields the highest cited papers belong to. They also pinpointed the sites where the best papers are delivered.

In his research, Muhammad Kamran et al. [25] presents a bibliometric analysis of articles in the Blockchain in the Internet of Things (BIoT) sector, encompassing papers published in prominent journals and conferences, and identifies research trends. It also looks at various study disciplines, the most prominent publications, the best publication venues, the best funding agencies, and the future of research. Our research will be quite similar to this, however our focus will be on data science.Missen et al. [26] and others give a case study of Pakistan to examine the impact on international research from 2009 to 2018. The study looks at 2000 articles written by 50 research scholars from various areas. This research is carried on three different levels: researcher, field, and domain. Readability scores, title formats, single and multiple authorships of papers, citation rates, publishing rates over time, the research contribution of both genders, and the impact of authors' Ph.D. institutions on research publications are all discussed in this work.

Fiala el al. [27] gives a bibliometric analysis of 1.9 million computer science papers indexed in Web of Science from 1945 to 2014. They examine the amount as well as the impact of these publications by document type, language, discipline, country, institution, and publication source. The most common keywords, cited references, and cited articles are also investigated, as well as the distribution of the number of references and citations per work and the age of cited references. They analyze the time and place of computer science conferences in terms of the most prolific months and locales because conference proceedings play such an important role in this scientific discipline. Finally, the production of journal articles and conference papers throughout the study,

Query
CU=*(PAKISTAN)* AND **SU**=*(Computer Science)*
Indexes = *SCI-EXPANDED, SSCI, A&HCI, ESCI*
Timespan = *(All years)*

TABLE 1. Query to get data from Web of Knowledge

as well as the level of collaboration within different computer science disciplines, are examined. One of the most important findings is that "Artificial Intelligence" is the most popular field in computer science.

Ding et al. [28] conducted a syntactic and semantic analysis of citations in computer science papers. The syntactic part entails identifying the location where the citations can be found (i.e. in which section of the article). Through a manual technique of predefined categorizations or an automated approach of NLP, semantic analysis determines the reason for citations. Citations are classified into categories defined by words or phrases in a decision tree in a predetermined classification.

III. DATA COLLECTION

The data on publications by Pakistani authors is the most essential part of our study. We gathered data from Web of Knowledge of Clarivate Analytics, commonly known as Web of Science. Web of Science is a subscription-based website that offers access to different citation databases. These databases include detailed citation information for a variety of academic areas. The Institute for Scientific Information created the Web of Sciences, which is now maintained by Clarivate Analytics. It has two search modes: basic search and advanced search. The advanced search option allows the users to build a query based on their needs and retrieve the required information. The data for publications by Pakistani researchers was generated using the query in table 1.

The abbreviated terms in table 1 represents **CU** as Country/Region, **SU** as Research Area, **SCI_EXPANDED** as Science Citation Index Expanded, **SSCI** as Social Science Citation Index, **A&HCI** as Arts & Humanities Citation Index, **ESCI** as Emerging Sources Citation Index. With a few exceptions in other languages, the above query revealed a total of 8,807 publications in English. The information was extracted and saved as a text (.txt) file. The data-set contains 8807 rows where each row represents single publication and the associated 60 columns represents different information about the publication.

A. TOOLS
The data in this study is analyzed using a variety of tools. In Jupyter Notebook, libraries such as NumPy, matplotlib, seaborn, pandas, and others are used to examine the data.

B. FEATURE SELECTION
We'll now go over the feature selection and preprocessing procedures we used to prepare the mined data for analysis. The data is extracted into 18 plain texts (.txt) files. The text

Feature Name
WR (Writer)
11 (Writer Full Name)
AF (Author Full Name)
TI (Document Title)
SO (Publication Name)
DT (Document Type)
AB (Abstract)
EM (E-mail Address)
NR (Cited 11 erence Count)
TC (Times Cited)
Z9 (Total Times Cited Count)
U1 (Usage Count (Last 180 Days)
U2 (Usage Count (Since 2013)
PU (Publisher)
PT (Publication Time)
P 11 ublished Year)
PG (Page Count)
WC (Web of Science Categories)
SC (Subject Category)
FU (Grant Number and Funding Agency)
RI (ResearcherID Number)

TABLE 2. Features used for Analysis

data files are then converted to Comma Separated Values (CSV) format using internet tools [29]. We merged these files into one data frame using the Pandas tool and Comma Separated Values (CSV) files. Then we looked at the different columns and found the key features that we might use in our analysis and study. We also looked at the columns that weren't relevant to our research and removed them from the dataset. After removing non-essential columns from the dataset, we were left with 22 columns on which we conducted the analysis.

The columns were mapped with truncated columns from the science scope website [29]. We now have 8804 records of articles and 22 columns indicating the various attributes of these articles after all of the data processing procedures. The majority of the data is made up of string columns, with a few float columns thrown in for good measure. The dataset also contains missing values. For a specific analysis, we dealt with these missing values in a unique method. The remaining characteristics are listed in Table No. 2.

C. PROCESS

This project's methodology consisted of gathering data, processing it, and extracting useful insights. Figure 1 explains it in more detail.

D. QUESTIONS

We have designed a questionnaire based on our data collection, which will provide relevant information and findings. Following are the research questions that serve as the foundation for our investigation and aid in the extraction of useful information.
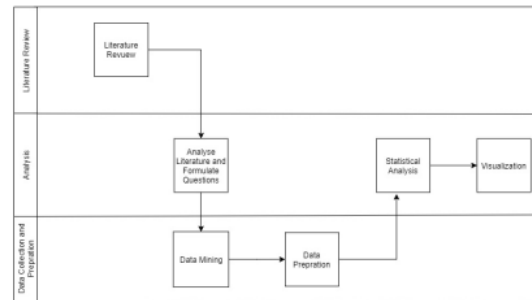

FIGURE 1. Methodology

1) What is the status of publications and citations in the computer sciences?
2) According to the number of publications, which study fields in Computer Science have the most research?
3) According to the number of citations, what are the most influential articles in the Computer Science domain?
4) What are the most popular venues for articles linked to computer science?
5) Which funding agencies are the most supportive of Computer Science papers?
6) What is the Computer Science domain's future research direction?
7) What are the most often mentioned publications?
8) What are how Pakistani researchers collaborate with researchers from other countries?
9) Which magazines were the most popular each year?
10) Determine how many publications have the highest number of citations across the board.
11) What is the average number of citations per publication?
12) What are the publications with several authors of one or more?

IV. EXPERIMENTS

We employed different bibliometric indices to assess the effect of research published by Pakistani authors in the field of computer science. The details of the bibliometric indicators we used are listed below.

A. ANALYSIS BASED ON METADATA

Count of Publications (**P**) per Author: How many articles did an author publish? Count of Publications (**P**) by Country: How many articles were published by each country? Count of References per Article: How many references were utilized in each article?

B. ANALYSIS BASED ON CONTENT

The most frequently used word in the Document Abstract, Author Keywords, and Document Title is represented via a word cloud.

C. ANALYSIS BASED ON CITATION

Average Citation Per Year: How many citations do you get on an annual basis? Total Number of Citations against a Keyword: Average Citation per Author.

D. ANALYSIS OF SOCIAL NETWORKS

When we can't see the relationships between entities and statistical analysis can't extract enough information, we employ social network analysis. The use of social network analysis to investigate the relationships between entities is quite useful.

E. LOTKA'S LAW

Lotka's law is one of the most often utilized bibliometrics laws for determining specific accomplishments in various **PAKISTAN JOURNAL AND PAPERS** of the Computer Science sector. Lotka's law is a discrete probability distribution function that reveals how productive an author is in a specific topic.

The frequency of publishing by authors in each discipline is described by Lotka's Law. "The number," it states (of authors) "Those who make n contributions contribute around $1/n2$ of those who make one, and the proportion of all contributors who make a single contribution is about 60 percent (Lotka 1926, cited in Potter 1988). This suggests that 60 percent of all authors in any subject will have only one publication, whereas 15 percent will have two (1/22 times.60). 7 percent of authors will publish three times (1/32 times.60), and so on. Only 6 percent of authors in a subject will create more than 10 publications, according to Lotka's Law of Scientific Productivity. Lotka's law can be accurate in general but not statistically correct when applied to vast quantities of literature over a lengthy period.

It's frequently used to predict how frequently authors will appear in an online catalog (Potter 1988).From 1984 to 2021, there were 12352 writers in the field of computer science. Because 7459 authors contributed to a single manuscript, the proportion of authors who contributed to a single paper was 0.604 (7459/12352). We may use Lotka's law to calculate the number of authors who contributed to two papers. 0.604 * 1/ 22 = 0.151 (number of authors contributing 2 papers). 0.604 * 1/ 32 = 0.067 (number of authors contributing 3 papers). The figure 2 is a graph of author productivity based on Lotka's Law.
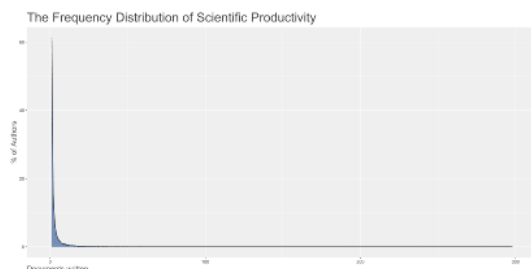
FIGURE 2. Lotka's Law

F. THE LAW OF BRADFORD

BRADFORD's Law is a rule of thumb for librarians when assessing the number of essential journals in a certain discipline. According to it, a single field's journals can be separated into three portions, each with the same amount of articles:

1) A small number of periodicals dedicated to the field that produce roughly one-third of all papers.
2) A second zone with the same amount of articles as the first but more journals.
3) A third zone, with the same amount of articles as the second zone but a higher number of journals

The relationship between the three zones was described by Bradford as 1:n:n2. Bradford's Law, as can be seen, is not statistically accurate in the strictest sense. However, it is still widely used as a rule of thumb (Potter 1988). The figure 3 shows that the core of journals (first zone) contains four journals: IEEE ACCESS (No. of documents = 1484), **INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS** (No. of documents = 719), and **INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS** (No. of documents = 719).**INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND NETWORK SECURITY** (Number of documents: 426) and **JOURNAL OF INTELLIGENT FUZZY SYSTEMS** (Number of documents: 299), with a total of 2928 documents in the journal's core. The second zone has 46 journals and a total of 3025 documents. The total number of journals in the third zone is 498, and the total number of papers is 2890.
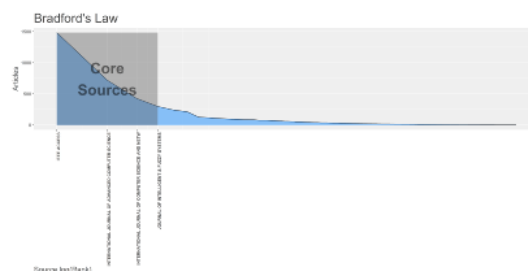
FIGURE 3. Bradfords Law

V. RESULTS

All of the analyses and findings of Pakistani Authors' Computer Science articles are presented and described in detail in this section. Because all of the findings are provided in depth in this section, the results are the most important phase of any research. The examination of the citation base is carried out in order to examine the research's impact. The study is based on publications. We did an analysis of the most popular publications of various types. This section also includes a content-based analysis in order to gain a better knowledge of research trends. The Effects of Various journals, publishers,

MAIN DESCRIPTION ABOUT DATA	
Documents	8843
Timespan	1984:2021
Sources (Journals, Books, etc)	548
Average citations per document	7.792
Average citations per year per doc	1.578
Average pages per document	13.902
References	268,817
Author's Keywords	25,515
Authors	12,352
Authors of single-authored documents	136
Authors of multi-authored documents	12,216
Single-authored documents	199
Documents per Author	0.716
Authors per Document	1.4
Co-Authors per Documents	4.5

TABLE 3. Features used for Analysis

and funding organisations are involved. This is also covered in this section.

A. SNEAK PEAK

The data from the Web of Science is shown in Table 3. Pakistani authors have written a total of 8843 documents in the field of computer science. These publications were published with a total of 548 sources, covering the years 1984 to 2021, with an average of 7.792 citations per document. Each paper has an average of 13.90 pages. There are also 268,817 references in the publications. The total number of Author Keywords utilised in the publications is 25,515, with a total of 12,352 Authors. There are only 136 single-authored documents with 136 authors and 12,216 multi-authored documents with 12,216 authors. The average number of authors per document is 1.4, while the number of documents per author is 0.716. Table No. 3 lists some of the other essential facts concerning data.

B. CITATION ANALYSIS

The number of citations a paper receives is an important factor in determining its value in the research world. The more often a document is quoted, the more respectable it and its contents become. We looked at the citation trend in this section.

1) **Average Article Citation per Year**

   Citation analysis can be used to determine the influence of works by demonstrating field usage and value. This can be a very useful metric for determining the annual effect of publications. In 2021, the maximum average citation per year attained is 3.0, as shown in Figure 4. From 1984 to 2012, the average number of citations per year was less than 2.0. According to the research, average citations have been rapidly growing since 2012.
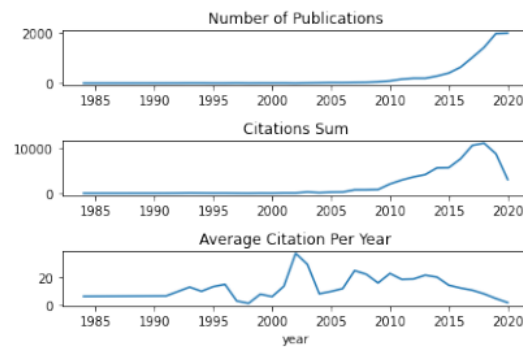

FIGURE 4. Average Article's Citation Per Year

2) **Citation Distribution**

   The analysis of citations is crucial for calculating the impact of research. By analysing the distributions of Citations, we were able to determine how often these research publications are used and referred by other authors. Figure 5 shows that around 38% of the documents only received citations ranging from 1 to 5. The most concerning statistic is that 28% of the documents received no citations, and only 3% of the documents received more than 50 citations.

3) **Most Cited Articles**

   The citations are ranked by the number of citations they received. The publications that have a higher number of citations are more useful and impactful. In table 4 lists the top ranking research articles by Pakistani writers in the field of computer science. The leading publication on the list is published, as can be seen.
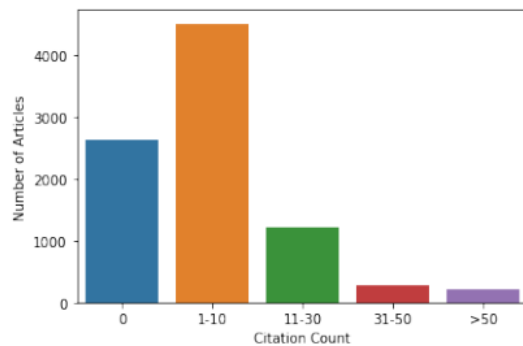

FIGURE 5. Citation Distribution Plot

in the Web of Science category Computer Science, Theory Methods by Khan, Minhaj Ahmad|Salah, Khaled in 2018. This paper's research focuses on IoT security, Block-chain, and IoT protocols. Cloud computing, wireless sensor networks, and artificial intelligence are among the other important study areas. In contrast, we discovered that publications published in

Publication Name	Pub Year	Citations	Funded
IoT security: Blockchain...	18	390	No
Soft sets combined...	2010	371	Yes
Mobile Cloud Comp...	2014	304	No
Security in cloud...	2015	291	Yes

TABLE 4. Most Cited Publications

2020 have the highest total number of citations, with 64. It's also worth noting that the top papers published in 2020 are medical-related, which is relevant to the global epidemic caused by COVID-19.

C. ANNUAL SCIENTIFIC PRODUCTION

A reliable metric of how well the scientific community is performing is the number of articles published each year. The graph depicts a comprehensive representation of papers from 1984 to 2021. The number of articles published per year was close to nil between 1984 and 2005. The number of published articles increased in 2005, indicating an upward trend. We can see from V-C Since 2010, the number of publications has been continuously increasing. For example, in 2010, the number of articles published per year was up to 250, and in 2019, we have over 1800. The trend in 2020 will continue to follow that of 2019.
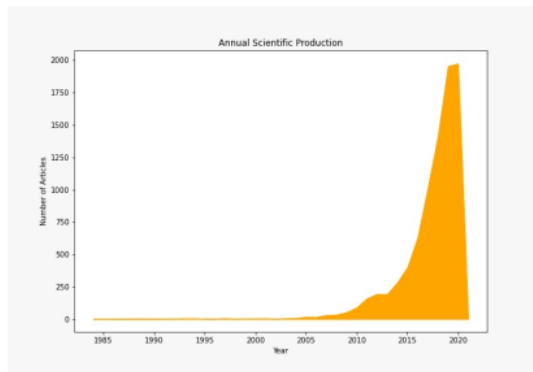

FIGURE 6. Annual Scientific Production

As shown in Figure V-C, the most productive years for research papers in the discipline of computer science were 2019 and 2020. As a result, we looked at the Monthly Production Trend for both years. Figure No. V-C depicts the research trend for the year 2019. We noticed that the beginning of 2019 was extremely productive, and the graph shows a general increase in yield, with the exception of February and August. In terms of the number of publications, the year 2020 was the most productive. We can observe in Figure No. V-C that the start of the year 2020 was quite productive, but the graph indicates a sharp decline after June. This pattern can be linked to the global impact of the COVID-19 epidemic. If there was no pandemic, the total number of publications may be higher than the total attained.
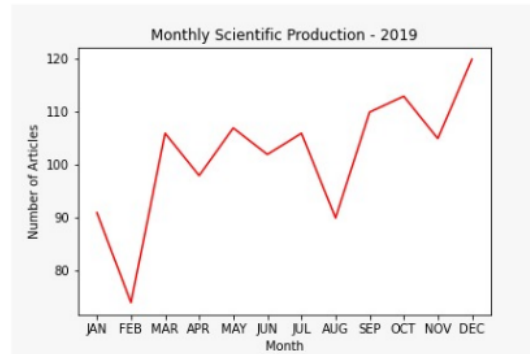
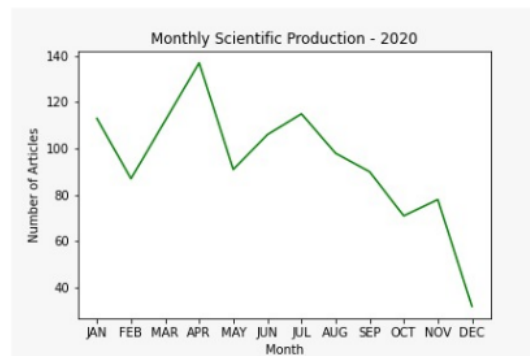**FIGURE 7.** Monthly Scientific Production - Year2019

FIGURE 8. Monthly Scientific Production -Year 2020

D. AUTHORS CORRESPONDING COUNTRY

The Web of Information is also used to do research on international by collaborating together calculating the number of publications published by the same country's authors vs papers by authors from different nations. To put it another way, there are two types of articles: single country publications (SCP) when all of the creators are from the same nation and different country publications (MCP) where the creators are from other countries.

The strongest literature creation and collaboration in the discipline of Computer Science is seen in Figure V-D. China, Korea, Malaysia, and Saudi Arabia have all contributed significantly to the development of literature and scientific research in Pakistan. Figure V-D further reveals that the vast majority of publications are published by authors from the same country. This could be because authors prefer to collaborate with people from their own study group or with academics from the same country. We may also see that Pakistani scholars work on very few projects with researchers from neighbouring nations such as India, Iran, and Afghanistan.

Figure V-I shows that the leading research financing entities in Pakistan are Chinese, Korean, and Saudi Arabian. We may deduce that Pakistan's high level of engagement with these countries is attributable to the country's well-
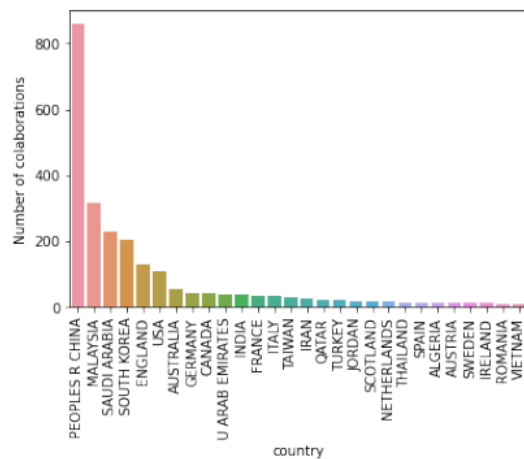
FIGURE 9. Authors Corresponding Countries

funded programmes. There are also a number of scholarship programmes between these nations, as well as a big number of Pakistani researchers obtaining PhDs in these countries. These types of programmes and collaborations are assisting Pakistani researchers in their efforts to advance and stay current in the research industry. These insights could aid us in identifying the best countries for collaboration.

E. KEYWORD GROWTH

The abstract is a summary of the publication's sections and provides a quick and easy way for active analysts to comprehend the paper. It could be a shortcut to the most important parts of your paper's written content. Many individuals only read the Abstract section of an article to get a general concept of what it is about. The following questions are addressed in the abstract section of the paper:

1) What was the outcome of the paper?
2) Why did we act in the way we did?
3) What conclusions have you reached?
4) Why are these discoveries significant and useful?

Keywords are devices that help indexers and search engines locate relevant publications. If databases and search engines are able to find your paper, users will be able to find it as well. This will increase the number of people interested in your paper, resulting in more citations.

1) The content of the paper is represented by keywords.
2) Keywords are unique to a certain field or sub-field.

The occurrences of words in the Abstract portion of the publications in our dataset are represented by Abstract Word Growth. Figure V-E depicts the increase in the number of words in the abstract section over time. Proposed, Data, and Paper are the most prevalent abstract words, with over 1000 occurrences every year, as seen in the graph. However, as shown in the graph, these abstract words became more prevalent after 2008. These abstract phrases are expected

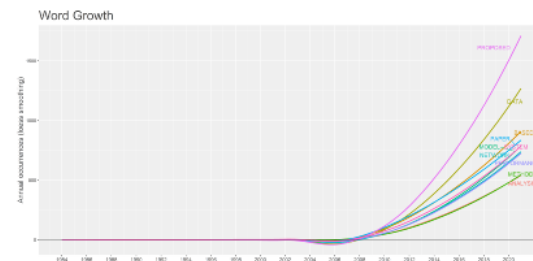to gain in popularity in the coming year, according to the projection.

FIGURE 10. Abstract Word Growth

Author Keywords are important terms or phrases that are utilised to locate information when researching a particular work. Keywords represent the paper's domain and are specific to the paper's field or sub-field of study. In Research Publication, after the Abstract, the author keywords are indicated by a KEYWORDS. The primary goal of keywords is to assist viewers in finding the paper in a list of related research publications. Keywords aid in the indexing of the work and improve its search-ability. As a result, it's critical for authors to include the most relevant keywords that will aid other authors in finding your research.
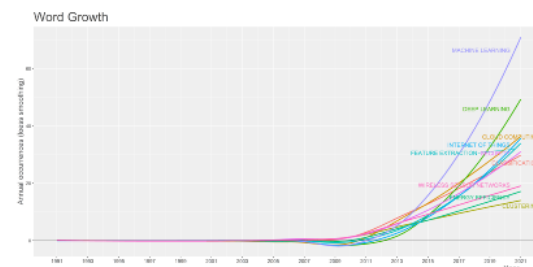

FIGURE 11. Author Keyword's Word Growth

The number of times a keyword appears in the Author Keyword section of the publications in our dataset is shown in Author Keyword Growth. The Author Keyword Growth graph, displayed in Figure No. V-E, depicts the evolution of the Author Keyword section's Keywords throughout time. It also shows which topic was investigated the most over time because keywords offer us an indication of the paper, thus we can deduce from the figure which topics and fields were studied. MACHINE LEARNING is the most common Author Keyword, with over 60 occurrences per year, as shown in the graph. MACHINE LEARNING is followed by DEEP LEARNING, which has around 50 occurrences per year. We might conclude from this fact that more study has been concentrated on Data Science.However, as shown in the graph, these abstract words began to gain popularity around 2012. These abstract phrases are expected to gain in popularity in the coming year, according to the projection. Another

observation is that the research of papers on themes related to CLUSTERING and ENERGY EFFICIENCY has slowed considerably throughout the years. The title of the paper is frequently the first impression that readers (and reviewers) get of it. As a result, a title should be attention-getting, accurately represent the contents of the paper, and entice people to read on. The title expresses the study's primary points and emphasises the significance of the research work. It is succinct and appealing to readers. The study paper's themes are listed in the title. The author tries to condense all of the subjects into as few words as possible in the title. The title of a publication piques other researchers' interest in reading and studying it.
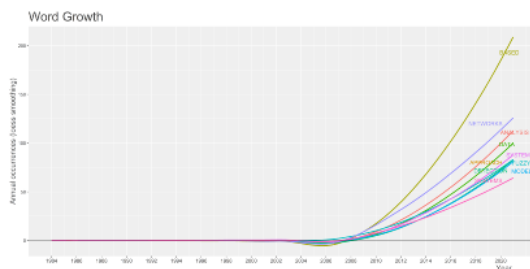

FIGURE 12. Title Word's Growth

The occurrences of keywords in the Document Title of the papers in our dataset are represented by Title Keyword Growth. Figure V-E shows the increase in the number of words in the Document Title over time. As seen in Figure V-E, BASED is the most commonly used title word, with over 200 instances per year. BASED is followed by NET-WORK, which has around 125 occurrences per year. Based on this observation, we can deduce that the majority of research document titles contain the words "based," "network," or "analysis." However, as shown in the graph, these words became more prevalent after 2008. These abstract phrases are expected to gain in popularity in the coming year, according to the projection.

F. WORD CLOUD ANALYSIS

Three word clouds are offered in this part for better visualisation of Bibliometric Analysis. The goal of using a word cloud for publications in the field of computer science is to assess the most common words that represent the majority of the work in that discipline. The trending themes in the field of computer science from 1984 to 2021 are depicted in Figure rV-F. The most important areas include deep learning, cloud computing, categorization, and feature extraction. On the list are also security, image processing, large data, and clustering.

The word cloud of the most prominent words in the field of computer science from 1984 to 2021 is shown in Figure No. V-F. The most important words in the title word cloud are networks, data, based, and analysis.

The abstract of a paper is a summary of the paper that includes the most important information. The abstract word
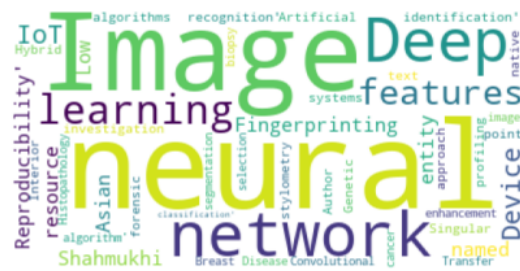

FIGURE 13. Author Keywords Word Cloud

FIGURE 14. Title Word Cloud

cloud is underneath image No. 16. The key phrases are proposed, network, systems, approaches, model, and algorithm.

FIGURE 15. Abstract Word Cloud

G. TOP JOURNALS

A journal is a scholarly periodical that includes articles authored by professors, researchers, and other specialists. Journals are devoted to a single discipline or subject of study. Journals, unlike newspapers and magazines, are written for academic or technical audiences rather than general readers. [?] The Institute of Electrical and Electronics Engineers Access (IEEE Access) is the top journal, with over 1400 publications, as shown in Figure 16. International Journal of Advanced Computer Science and Applications, as well as International Journal of Computer Science and Network Security, are promising publications with a large number of papers written by Pakistani authors. The most often cited journals are International Journal of Distributed Sensor Networks and Computer and Electrical Engineering Journal.

This also demonstrates the low level of research conducted by Pakistani scholars in the fields of sensor networks and electrical engineering.The Institute of Electrical and Electronics Engineers Access (IEEE Access) is the top journal, with more than 1400 publications, as shown in Figure V-G. International Journal of Advanced Computer Science and Applications and International Journal of Computer Science and Network Security are two other promising journals where Pakistani authors publish the majority of the papers. The most frequently denied journals are the International Journal of Distributed Sensor Networks and the Computer and Electrical Engineering Journal. This also demonstrates the low level of study conducted by Pakistani scholars in the areas of sensor networks and electrical engineering.
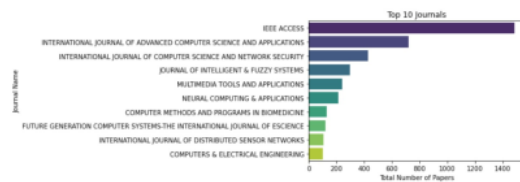

FIGURE 16. Top 10 Journals

The Source Growth graph in Figure V-G depicts the evolution of journal sources throughout time. IEEE Access, as shown on the graph, is the most popular journal, with approximately 350 occurrences per year. However, according to the graph, the source gained popularity after 2010, whereas before that, the only most popular sources were International Journal of Advanced Computer Science and Applications and International Journal of Computer Science and Network Security, with annual occurrences of about 100 and 90, respectively. Multimedia Tools and Applications, with a publication of around 40, has the lowest growth potential exhibited up to 2020.Furthermore, the journal International Journal of Advanced Computer Science and Applications shows no substantial improvement, and all others follow the same pattern. IEEE Access is leading Journal and will be leading in future years.
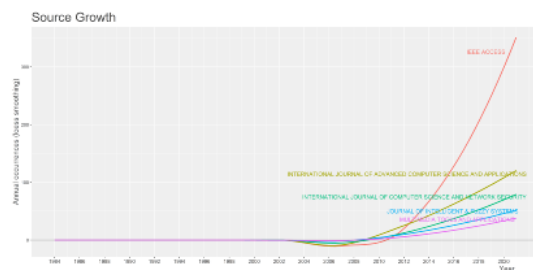

FIGURE 17. Source Growth

H. TOP PUBLISHERS
The entity that published the research papers is known as a publisher. Publishers publish research in different publications for distinct types of study. The IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC is the top publisher, with roughly 1750 publications, as shown in Figure No. V-H. SPRINGER is also the largest publisher, with over 1000 titles in print. The publishers with the most rejections are ELSEVIER SCIENCE BV and WILY.
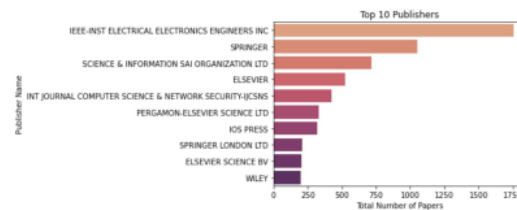

FIGURE 18. Top Publishers

I. TOP FUNDING ORGANIZATIONS
Funding is a significant aspect that influences the quality of research. The most focused and significant projects are usually those that are sponsored. Normally, funding agencies pay research organisations for high-end study on a certain issue. The National Natural Science Foundation of China (NNS-FOC) has supported the majority of the funded researches, as shown in Figure No V-I. This donation by a Chinese entity has elevated China to the position of Top Corresponding Author, as shown in the graph. The Higher Education Commission (HEC) has supported over 480 research projects and is currently in first place. Korea's National Research Foundation and Saudi Arabia's King Saud University are the biggest donors in Pakistani research, making these countries the Top Corresponding Authors Countries.
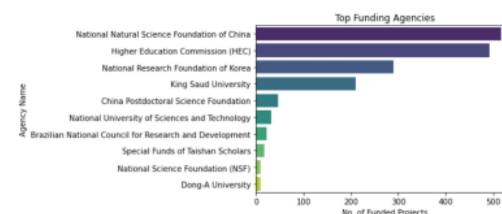

FIGURE 19. Top Funding Agencies

VI. CONCLUSION
From 1984 to 2021, this study presents a Bibliometric Analysis of Pakistani researchers' research output in the sphere of information technology. We discovered that the top funding agencies in terms of funding projects related to computer science research were the National Natural Science Foundation of China with 500 funded projects, HEC with more than 450 funded projects, National Research Foundation of Korea with 300 funded projects, and King Saud University, Saudi Arabia with nearly 250 funded projects. In Author Corresponding Country, we discovered both Single Country Publications

(SCP) and Multi-Country Publications (MCP).

We calculated the Annual Scientific Production based on the number of publications published throughout time, as well as the Citation Distribution based on the number of articles published. The co-occurrence network of Author Keyword, Title Word Growth, Abstract Word Growth, and Author Keyword is also discovered. In the Computer Science sector, Bradford's Law is used to figure out how many core journals and articles there are. In the field of computer science, Lotka's Law is used to characterise the frequency with which authors publish.

The amount of publications and citations in an area are heavily influenced by the most popular Author Keywords. We've noticed that data science-related subjects like Machine Learning, Deep Learning, and Cloud Computing have been trending. We also predict that in the next years, interest in Data Science will continue to expand.

We also identified Top Authors based on their articles over time. With over 200 articles published in 2020, KHAN. M is the top author. As a result, KHAN.M is the most productive author in terms of publications. Citation analysis is used to assess the influence of works by determining their field usage and worth. This can be a very useful metric for determining the impact of articles on an annual basis. Citation Analysis tells us how many citations there are for a certain keyword. The semantic relationship was extracted using Network Graph Analysis. The semantic relationship between the author keywords is extracted using the co-occurrence network graph.

For better visualisation of Bibliometric Analysis information, Word Cloud Analysis is used. The goal of using a word cloud in this paper was to examine the most common words that represent the majority of the work done in particular domains. The number of publications per author and the number of publications per country were calculated using Meta-Based, and the author's n-index indicates that n of the researcher's articles received n citations, as well as the number of references per article. The most frequently used word in the Document Abstract, Author Keywords, and Document Title was determined using a Word Cloud for Content Based Analysis.

REFERENCES

[1] Dodson, M. V. "Research paper citation record keeping: It is not for wimps." Journal of animal science 86.10 (2008): 2795-2796.

[2] Thomson Reuters. History of citation indexing. Essay in Free Scientific Resources.

[3] Garfield, Eugene. "Citation indexes for science. A new dimension in documentation through association of ideas." International journal of epidemiology 35.5 (2006): 1123-1127.

[4] Garfield, Eugene. "Citation frequency as a measure of research activity and performance." Essays of an Information Scientist 1.2 (1973): 406-408.

[5] Garfield, Eugene. "The use of journal impact factors and citation analysis for evaluation of science." 41st Annual Meeting of the Council of Biology Editors, Salt Lake City, UT. 1998.

[6] Canese, Kathi, and Sarah Weis. "PubMed: the bibliographic database." The NCBI Handbook 2 (2013): 1.

[7] Dittmar, Paul G., Robert E. Stobaugh, and Charles E. Watson. "The chemical abstracts service chemical registry system. I. General design." Journal of Chemical Information and Computer Sciences 16.2 (1976): 111-121.

[8] TePaske-King, Bert, and Norman Richert. "The Identification of Authors in the Mathematical Reviews Database." Issues in Science and Technology Librarianship 31 (2001).

[9] Bergmark, Donna, Paradee Phempoonpanich, and Shumin Zhao. "Scraping the ACM digital library." ACM SIGIR Forum. Vol. 35. No. 2. New York, NY, USA: ACM, 2001.

[10] Caragea, Cornelia, et al. "Citeseer x: A scholarly big dataset." European Conference on Information Retrieval. Springer, Cham, 2014.

[11] Analytics, Clarivate. "Web of science." (2017).

[12] Burnham, Judy F. "Scopus database: a review." Biomedical digital libraries 3.1 (2006): 1-8.

[13] Mongeon, Philippe, and Adèle Paul-Hus. "The journal coverage of Web of Science and Scopus: a comparative analysis." Scientometrics 106.1 (2016): 213-228.

[14] Falagas, Matthew E., et al. "Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses." The FASEB journal 22.2 (2008): 338-342.

[15] Jacsó, Péter. "Google Scholar: the pros and the cons." Online information review (2005).

[16] Waltman, Ludo, and Vincent A. Traag. "Use of the journal impact factor for assessing individual articles need not be statistically wrong." F1000Research 9 (2020).

[17] Curry, Stephen. "Let's move beyond the rhetoric: it's time to change how we judge research." Nature 554.7690 (2018): 147-148.

[18] Lowry, Paul Benjamin, et al. "Evaluating journal quality and the association for information systems senior scholars' journal basket via bibliometric measures: Do expert journal assessments add value?." MIS quarterly (2013): 993-1012.

[19] da Silva, Jaime A. Teixeira, and Aamir Raoof Memon. "CiteScore: A cite for sore eyes, or a valuable, transparent metric?." Scientometrics 111.1 (2017): 553-556.

[20] Mañana-Rodríguez, Jorge. "A critical review of SCImago journal & country rank." Research evaluation 24.4 (2015): 343-354.

[21] Iqbal, Waleed, et al. "A bibliometric analysis of publications in computer networking research." Scientometrics 119.2 (2019): 1121-1155.

[22] Ali, Muhammad Yousuf, and Joanna Richardson. "Research publishing by library and information science scholars in Pakistan: A bibliometric analysis." Journal of Information Science Theory and Practice 4.1 (2016): 6-20.

[23] Bajwa, Rizwan S., and K. Yaldram. "Bibliometric analysis of biotechnology research in Pakistan." Scientometrics 95.2 (2013): 529-540.

[24] Garousi, Vahid, and João M. Fernandes. "Highly-cited papers in software engineering: The top-100." Information and Software Technology 71 (2016): 108-128.

[25] Kamran, Muhammad, et al. "Blockchain and Internet of Things: A bibliometric study." Computers & Electrical Engineering 81 (2020): 106525.

[26] Missen, Malik Muhammad Saad, et al. "Scientometric analysis of social science and science disciplines in a developing nation: a case study of Pakistan in the last decade." Scientometrics 123.1 (2020): 113-142.

[27] Fiala, Dalibor, and Gabriel Tutoky. "Computer science papers in Web of Science: A bibliometric analysis." Publications 5.4 (2017): 23.

[28] Ding, Ying, et al. "Content-based citation analysis: The next generation of citation analysis." Journal of the association for information science and technology 65.9 (2014): 1820-1833.

[29] http://medialab.github.io/sciencescape/

[30] http://https://www.uvic.ca/library/research/tips/journal/index.php

• • •

Group_18_Project_report.pdf

| 10 | jistap.koar.kr
Internet Source | 1% |

| 11 | www.thefreelibrary.com
Internet Source | 1% |

| 12 | Malik Muhammad Saad Missen, Sajeeha Qureshi, Nadeem Salamat, Nadeem Akhtar, Hina Asmat, Mickaël Coustaty, V. B. Surya Prasath. "Scientometric analysis of social science and science disciplines in a developing nation: a case study of Pakistan in the last decade", Scientometrics, 2020
Publication | 1% |

| 13 | Samuel Fosso Wamba. "Humanitarian supply chain: a bibliometric analysis and future research directions", Annals of Operations Research, 2020
Publication | 1% |

| 14 | academic-accelerator.com
Internet Source | 1% |

Exclude quotes   On   Exclude matches   < 1%
Exclude bibliography   On