Digital Object Identifier

Bibliometric Analysis of Research Publications in Domain of Computer Science : Case Study of Pakistani Authors

HAMZA MUNIR[1], AGHA MUHAMMAD ANAS[2], AND MEHMOOD ALI.[3]
[1]Information Technology University, Lahore 54000 (e-mail: msds20061@itu.edu.pk)
[2]Information Technology University, Lahore 54000 (e-mail: msds20068@itu.edu.pk)
[3]Information Technology University, Lahore 54000 (e-mail: msds20090@itu.edu.pk)

Corresponding Author: Hamza Munir (e-mail: msds20061@itu.edupk).

ABSTRACT Bibliometric Analysis is the study of analyzing the scientific research patterns and trends in specific community, country and field etc. In this study, the bibliometric analysis of all computer science related publications by Pakistani authors is performed. The data is mined from web of science multiple Indexes. The numerous parameters are analyzed including Annual Scientific Production, Average Citation per Article, Most Cited documents, Most Relevant Authors, most relevant sources(journals) and Publishers, hot topics etc. The content-base, citation-base and time series based analysis is performed for better understanding of publications trends. The research pace in the past few years was very slow, whereas, from 2014-2020, it increases exponentially. IEEE Access is the most relevant journal, with 1400+ publications. The most corresponding authors are from China, Korea, Malaysia, and Saudi Arabia. The National Natural Science Foundation of China, Higher Education Commission and King Saud University are the top funding organizations for research. Machine Learning, Deep Learning and Cloud Computing are the most used author keywords and are hot fields in the recent years. The majority of the papers are Article type following by Review and Article; Early Access. The Average article citation was very low before 2004, After 2004 it increases by 50 percent. Whereas, more than 38 percent of the research publications get less than equal to 5 citations. 28 percent of the publications didn't get any citation and only 3 percent of the publications get more than 50 citations.

INDEX TERMS Bibliometric Study, Computer Science, Pakistani Authors ,Research Patterns, Citation Analysis

I. INTRODUCTION

Bibliometric Analysis is used for research assessment worldwide. Bibliometric Analysis is the strong methodology to analyze the trends in research publications for better planning of future research strategies. Bibliometric Analysis is a technical skill and great Critical thinking and a research knowledge is required to interpret bibliometric data correctly.

Bibliometric analysis of a publications is a crucially important source of objective knowledge and information about the quantity and quality of scientific work(Narin et al. 1994). In this research, we did a bibliometric analysis of the publications of the field of computer science by Pakistani authors. Research in computer science is a exponentially growing research domain in this century. Our research helps the Pakistani researchers to easily identify the trending topics in the domain of computer science. This study will help them to choose research topics for future.

There are three publication types in our data, such as Journals, Series and Books. To keep the scope of this study to manageable proportions, we have focused on journal publications. We have analyzed the Top Publishers and Top Journals where Pakistani researchers published their papers. The Top Publishers were IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC, SPRINGER, SCIENCE & INFORMATION SAI ORGANIZATION LTD etc. The Top journals were IEEE ACCESS, INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, INTERNATIONAL JOURNAL OF COMPUTER

SCIENCE AND NETWORK SECURITY, JOURNAL OF INTELLIGENT & FUZZY SYSTEMS etc. We analyzed the research trends with respect to year using different statistical techniques. We also analyze popular topics in Journal on computer science.

We investigate changes in publication behavior over the years. We perform the content base analysis, The detailed analysis of keywords used in document's title, abstract and author keywords is performed. We analyzed the collaboration of Pakistani researchers with other countries. We also made some questions in mind by seeing data available which helps us to keep the focus of research. Our research answer's the following questions: Which topics are trending in domain of computer science? How research activities are improving over time? The key contribution of this article is to do a detailed analysis on Pakistani researcher contribution to the world in domain of computer science.

The rest of this article is structured as follows. In section 2, we discussed about the related research work in the domain of computer science and related domains. In Section 3, our dataset, dataset preparation, tools, feature selection and methodology is explained. . In section 4, our Results are broadly explained with multiple graphs to validate our reasoning. In section 5, we conclude our findings.

II. LITERATURE REVIEW

IN this section we take a look at the previous work done for bibliometric analysis in different fields of science. The purpose is to take a general look at the different approaches adopted by different researchers to present their work in an efficient way possible.

Iqbal, W., Qadir, J., Tyson, G. et al [1] analyses the computer networking domain and gives a study which articulates content and metadata of four important computer networking periodicals —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)—obtained using ACM, IEEE Xplore, Scopus and CrossRef, for an 18-year period (2000–2017) to address important bibliometric questions. In the course of this study, they tried to estimate the co-existence of trends in the COMST and TON journals and check their comparison with the publication trends in INFOCOM and SIGCOMM. After the comparison of journals with the publication trends the analyses of the computer networking literature comprises the following: (a) metadata analysis; (b) content-based analysis; and (c) citation analysis. In addition, using the number of publication and article citations. they identified the most impactful trends and the most dominant authors, institutes and countries.

Ali, Muhammad & Richardson, Joanna [2] performed the bibliometric analysis of the research published by the LIS (Library and Information Sciences) scholars in Pakistan through a survey. A questionnaire was designed by the authors and distributed to the LIS scholars through Email, Yahoo groups and Facebook to the representatives from all the provinces of Pakistan. The results obtained from the total of 104 respondents were then analyzed in the tool SPSS version 21. They pointed out the significant demographic information, for example gender and geographical, the degree of shared author-ship, the range of publishing based on geological regions, the stability of relationship between job title (seniority) and publication count, the stability of citation metrics for national output, recognizing the factors which may adversely affect on LIS researcher's capacity to take research in its undertaking.

R. S. Bajwa and K. Yaldram [3] presents an analysis of the research trends in Pakistan in the field of biotechnology for the period 1980–2011. They analyzed the increase and decrease of Annual growth rate, comparison of organization who actively participate in research in biotechnology using Publication rate, citation rate average citation per paper and multiple indexing methods.

Garousi and Fernandes [4] computed and classified the top 100 papers in the field of software engineering. In their study they found the top papers based on the citation count and the average annual number of citations. They devised a GQM (Goal, Question, and Metric) methodology to find their goal by devising some questions in order to achieve it. They have also compared the top papers with the top papers of all areas of science. They also identified the fields to which the top cited papers are related. Further they also identified the venues in which the top papers are presented.

Muhammad Kamran, Hikmat Ullah Khan [5] in research presents a bibliometric examination of articles in Blockchain in Internet of Things (BIoT) subject, covering papers published in the trending journals and conferences, and discovers research patterns. It also investigates different research regions, the most compelling publication, top publication venues and future examination direction. The research we are going to conduct is very similar to this but our area of interest is computer science including all domains.

Missen, M.M.S., Qureshi, S., Salamat [6] and other presents Pakistan case study to analyze the impact on International research from the timespan 2009 to 2018. The study inspects 2000 published articles by 50 researchers of various disciplines. This investigation is directed on three different levels: scientist level, field level and domain level. In this paper they examine readability scores, title formats, single and different authorships of articles, reference citation rates, publication rates over the long run, the research commitment of both genders and the effect of PhD institutions of authors in research publications.

Fiala, Dalibor & Tutoky, Gabriel [7] presents a bibliometric analysis of 1.9 million computer science papers published from 1945 to 2014 and ordered in the Web of Science. They examine both the amount and the effect of these publications as indicated by document types, dialects, disciplines, countries, institutes, and publication sources. The most regular authors keywords, referred citations, and referred to papers just as the dispersion of the quantity of references and ref-

erences per paper and of the period of referred to references are likewise investigated. As conference procedures assume a gigantic part in this logical field, they examine the time and place of computer science conferences as the most productive months and areas. The long and short of the study is the creation of journal articles and conference papers throughout the entire time span and the degree of joint effort in various computer science disciplines are assessed. One of the fundamental outcomes is the finding that "Artificial Intelligence" is the most profitable subfield of computer science, “Neural Networks” [10] seems to be losing its charm but “Interdisciplinary Applications” has the highest relative impact.

In another research Bakri, A., & Willett, P. [8] analyses the computer science publications in Malaysia. The motivation behind this paper is to examine the publications of, and the references to, the present staff of 19 departments of computer science in Malaysian universities, to compare these bibliometric data with expert peer reviews of Malaysian research performance. This examination look through reference of the Scopus and Web of Science data sets. The decision make from research is both publications and reference rates are low, although this is at any rate to some degree because of some Malaysian universities having just an instructing capacity. A greater amount of the department’ publications were distinguished in Scopus than in Web of Science, yet the two databases were required for far reaching inclusion. Statistically significant relationships were seen between the departments’ publications and reference checks and the rankings of the departments’ parent universities in two assessments of the research performance of Malaysian universities.

Ding et al. [9] performed the syntactic and semantic analysis of the citations in publications related to computer science. The syntactic part involves the identification of the location where the citations are found (i.e. in which section of the article). The semantic analysis finds the motivation of the citations through manual approach of predefined categorizations or semiautomatic approach of Natural Language Processing. In predefined categorization the citations are divided into categories defined by words or phrases in a decision tree.

III. DATA COLLECTION AND METHODOLOGY

We start by describing our data collection process and methodology. There are several types of document genres in the field of computer science, including articles, editorials, books, survey papers, and studies. Multiple data columns are analyzed in detail to mine useful Information.

A. DATA COLLECTION

To perform the analysis, data is collected from Web of Knowledge also known as Web of Science, Clarivate Analytics[11]. Web of Science is a website that provides subscription-based access to multiple databases that provide comprehensive citation data for many different academic disciplines. It was originally produced by the Institute for Scientific Information and is currently maintained by Clarivate Analytics.[13] Using Web of Knowledge website[11], from advanced search, data is queried using input parameters in TABLE 1.

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

The 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.

The query mentioned above returned a total of 8,843 publications with publications in English with a few exceptions in other languages. The results were extracted and stored in the form of Text(.txt) format file. The data was downloaded on 18th December 2020. The data-set contains 8843 rows, each row represents single publication and 60 columns representing different information about publications.

B. TOOLS

In this research, multiple tools are used to analyze the data. The Python programming language in Jupyter Notebook is used to analyze the data with libraries such as numpy, matplotlib, seaborn, pandas etc. In parallel, we use R language in R studio with libraries Bibliometrix and Biblioshiny to get advance level visualizations.

C. FEATURE SELECTION

We now describe the feature selection and preprocessing steps we performed on the mined data for further analysis.

The dataset is mined in the 18 plain text (.txt) format files. We then use online tools to convert the text data files into Comma Separated values (CSV) format[12]. Using Pandas tool and Comma Separated values (CSV) files we merged these files and make 1 dataframe. We then analyze the different columns and identified the important feature which can be useful for our analysis and research. We also analysed the columns which cannot be useful for our research and remove them from dataset. After dropping non important columns from dataset we left with 22 columns on which we performed the analysis. The dataset contains author details as well as publication details such as author’s name, publication name, citation count, publication year, publisher name, references and many more important features.

In Data cleansing process, irrelevant data is removed from the dataset like features, entries without metadata such as author names, columns with very few values. For different analysis multiple new features are also created to support our reasoning. The major issue with columns of data is that they were represented by Short forms of column names. To solve this problem, we use Beautiful Soup to scrap the columns

Feature Name
AU (Authors)
AF (Author Full Name)
TI (Document Title)
SO (Publication Name)
DT (Document Type)
DE (Author Keywords)
AB (Abstract)
C1 (Author Address)
EM (E-mail Address)
NR (Cited Reference Count)
TC (Times Cited)
Z9 (Total Times Cited Count)
U1 (Usage Count (Last 180 Days)
U2 (Usage Count (Since 2013)
PU (Publisher)
PD (Publication Date)
PY (Year Published)
PG (Page Count)
WC (Web of Science Categories)
SC (Subject Category)
FU (Funding Agency and Grant Number)
RI (ResearcherID Number)

TABLE 2: Features used for Analysis

mapping from sciencescape website and mapped the columns with abbreviated columns.[12]

After all the data processing steps, we have now 8804 records of articles and 22 columns representing the different features of these articles. Majority of the data contains the string type columns and few of them are float type. There are also missing values in the dataset. We handled these missing values in specific way for specific analysis. The remaining feature are mentioned in the Table No. 2.

D. PROCESS

The methodology we have used for this project is getting data, process it and extract fruitful insights. It is further explained in Figure 1 Given Below:

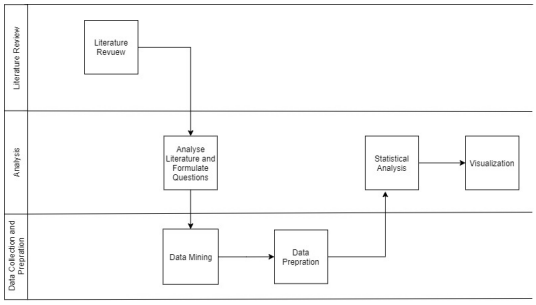

FIGURE 1: Methodology

E. QUESTIONS

We also created a questionnaire according to our data set that will provide useful results and information. In the following, we present the research questions that form the basis of our research and help us to extract fruitful insights.

1) What is the trend of Computer Science publications and citations?
2) Which research areas have more Computer Science research based on the number of publications?
3) What are the most influential papers in Computer Science domain according to the number of citations?
4) What are the most popular publication venues for Computer Science related papers?
5) What are the top most supportive funding agencies of Computer Science papers?
6) What is the future research direction in the Computer Science domain?
7) Determine the top cited publications?
8) Determine how Pakistani researcher are collaborating with other countries?
9) Determine which publications were top for each year?
10) Determine how many publications have recorded maximum citations throughout?
11) Determine average citations per Publication?
12) Determine the publications where Number of Authors are one and greater?
13) Determine average number of authors per publications?

F. EXPERIMENTS

In order to measure the impact of research published by Pakistani authors in domain of computer science, we used several bibliometric indicators. Following are the details of bibliometric indicators we used.

A. META-BASED ANALYSIS Count Of Publications (P) per Author : Author published how many number of Articles. Count Of Publications (P) per Country : Country contributed in publishing how many number of Articles. Count Of References per article : Article used how many number of references.

B. CONTENT-BASED ANALYSIS Word Cloud : Most frequent word used in Document Abstract, Author Keywords and Document Title.

C. CITATION-BASED ANALYSIS Average Citation per Year : What is the average citations per year. Average Citation per Author : Total Number of Citations present against a Keyword.

D. SOCIAL NETWORK ANALYSIS Social Network Analysis is used when we cannot observe relationship between entities and cannot extract enough information through statistical analysis. Social Network Analysis is great way to explore the relation between entities.

We also used most commonly used laws in bibliometrics in our research: Lotka's law of scientific productivity and Bradford's

E. LOTKA'S LAW Lotka's law holds a very important place amongst the most frequently used laws of Bibliometrics to determine certain accomplishments in various PAKISTAN JOURNAL AND PAPERS of Computer Science field. Lotka's law is a discrete probability distribution function that tells the author productivity within that domain.

Lotka's Law describes the frequency of publication by authors in each field. It states that "the number (of authors)

making n contributions is about 1/n² of those making one; and the proportion of all contributors, that make a single contribution, is about 60 percent" (Lotka 1926, cited in Potter 1988). This means that out of all the authors in each field, 60 percent will have just one publication, and 15 percent will have two publications (1/2² times .60). 7 percent of authors will have three publications (1/3² times .60), and so on. According to Lotka's Law of scientific productivity, only six percent of the authors in a field will produce more than 10 articles. Lotka's Law, when applied to large bodies of literature over a fairly long period of time, can be accurate in general, but not statistically exact. It is often used to estimate the frequency with which authors will appear in an online catalog (Potter 1988).

In the field of computer science from year 1984 to 2021 we have 12352 authors. The number of authors which contributed to one paper were 7459, so proportion of author who contributed one paper was 0.604 (7459/12352). And to calculate number of authors who contributed 2 paper we can apply here Lotka's law. No. of author contributing 2 paper = 0.604 * 1/ 2² = 0.151. No. of author contributing 3 paper = 0.604 * 1/ 3² = 0.067. Below is the figure of author productivity by using Lotka's Law:

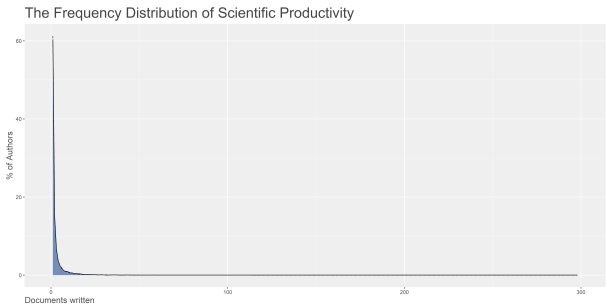

FIGURE 2: Lotka's Law

F. BRADFORD'S LAW Bradford's Law serves as a general guideline to librarians in determining the number of core journals in any given field. It states that journals in a single field can be divided into three parts, each containing the same number of articles:

1) 1) A core of journals on the subject, relatively few in number, that produces approximately one-third of all the articles.
2) 2) A second zone, containing the same number of articles as the first, but a greater number of journals.
3) 3) A third zone, containing the same number of articles as the second, but a still greater number of journals.

Bradford expressed the relationship between 3 zone as 1:n: n². As may be seen, Bradford's Law is not statistically accurate, strictly speaking. But it is still commonly used as a general rule of thumb (Potter 1988). In the figure below we can see that core of journals (first zone) contain 4 journals, IEEE ACCESS (No. of documents = 1484), INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS (No. of documents = 719), INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND NETWORK SECURITY (No. of documents = 426) and JOURNAL OF INTELLIGENT & FUZZY SYSTEMS (No. of documents = 299) and total documents in core of journal are 2928. Second zone contain 46 journal and total no. of documents are 3025. In third zone total journals are 498 and total no. of documents are 2890.

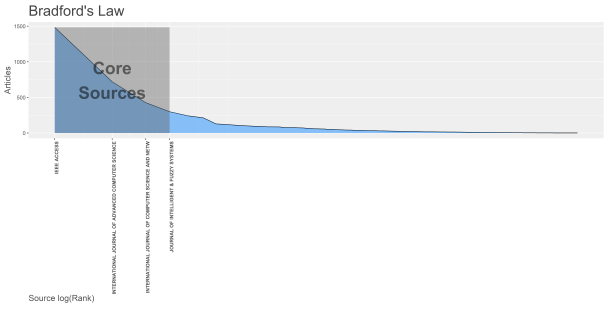

FIGURE 3: Bradfords Law

IV. RESULTS

In this section, All the analysis and finding about the Computer Science publications by Pakistani Authors are presented and explained in details. Results are the key phase of any research because all the findings are presented in this section in details. The Citation base analysis is performed to analyse the impact of research. Publication base analysis is performed to analyze the top publications. Different types of content base analysis is also performed in this section for better understanding of research patterns. The Impact of Different journals, Publishers and Funding Organization are also discussed in this section.

A. SNEAK PEAK

Table.3 shows the data extracted from the Web of Science. A total of 8843 documents are present in the field of Computer Science by Pakistani Authors. These documents published with a total of 548 sources, the period of all documents is 1984 to 2021, and 7.792 is the average citations per document. The average number of pages per document is 13.90. Moreover there are 268,817 references used in papers. Author Keywords used in the documents are 25,515 by total 12,352 Authors. There are only 136 Authors of single author documents and 12,216 authors of multi-authored documents. The average author per document is 1.4 and documents per author rate is 0.716. Some of the other important statistics about data are mentioned in Table No. 3.

B. CITATION ANALYSIS

The number of citations of a document plays a considerable role in increasing its worth in the scientific community. The more widely cited the document is, the more prestigious it becomes along with its contents. In this section we have analyzed the trend of citations.

MAIN DESCRIPTION ABOUT DATA	
DESCRIPTION	RESULTS
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: Basic Statistics about Data

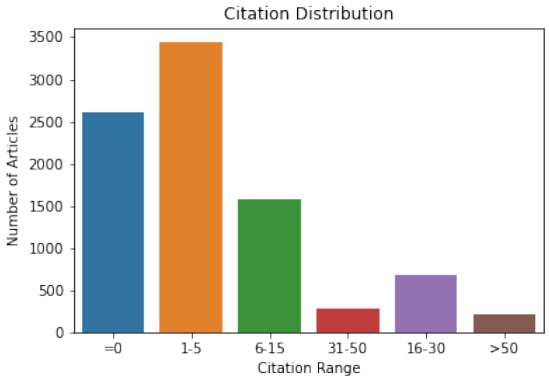

FIGURE 5: Citation Distribution Plot

1) Average Article Citation per Year

Citation analysis can used to measure the impact of papers, indicating field usage and value carried by that article. This can be very useful measure to assess the yearly impact of articles. In the Figure No. 4, we can see that the maximum average citation per year achieved is 3.0 in 2021. From 1984 to 2012 average citation per year was below 2.0. The study reveals that average citation shows rapid increasing trend after the year 2012.

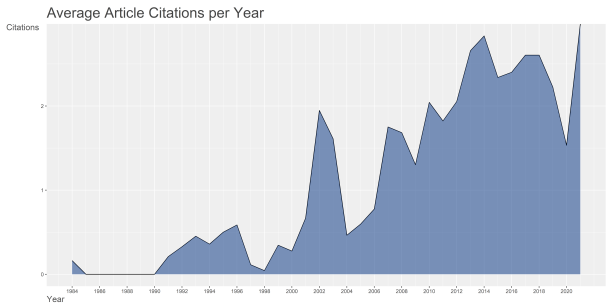

FIGURE 4: Average Article's Citation Per Year

2) Citation Distribution

To calculate the impact of research, The analysis of citations is very important. We analyzed that how much these research publications are used and referenced by other Authors by making the distributions of Citations. From Figure No. 5, we can see that the around 38 percent of the documents only got citations from 1-5. The most alarming thing is the 28 percent of the documents didn't get any citation and Only 3 percent of the documents got more than 50 citations.

3) Most Cited Articles

The citation are ranked on the basis of how much citation it got. The papers wit more citation are more useful and impactful. The top ranked research articles in the domain of computer science by Pakistani authors are listed in Table No. 4. We can see that the leading publication in list is published in 2018 by Khan, Minhaj Ahmad|Salah, Khaled in the Web of Science category Computer Science, Theory & Methods. The research focus in this paper is about IoT security, Block-chain and IoT protocols. The other leading research publication are from the cloud Computing, Wireless sensor networks and Artificial Intelligence. Whereas in detailed research we found the papers published in 2020 has got the maximum total citation are 64. It is also observed that the top articles published in 2020 are focused on medical which we can relate with global pandemic due to COVID-19.

Publication Name	Pub Year	Citations	Funded
IoT security: Blockchain...	2018	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

C. ANNUAL SCIENTIFIC PRODUCTION

A number of articles published per year are a reliable indicator of how well the scientific community is performing. The Figure provides a thorough visualization of articles from 1984 up to 2021. From 1984 to 2005 the number of articles published per year was close to zero. In the year 2005, the number of published articles shows the increasing trend. From Figure No. 6, we observed after 2010 the number of articles has been increasing significantly over the years. For Instance, in 2010, the numbers were up to 250 articles a year, and in the year 2019, we have more than 1800 articles. In 2020 the trend keeps on following the 2019's trend.

As Figure No. 6 shows, 2019 and 2020 were the most productive years for research publications in domain of computer science. So we analyzed the Monthly production trend of publication for both years. Figure No. 7 Shows the research trend of year 2019, we observed that the Start of 2019 was very productive and the graph shows overall increasing trend except for the months of February and August.

The year 2020 was most productive year in terms of the Number of Publications. From Figure No. 8, we can see

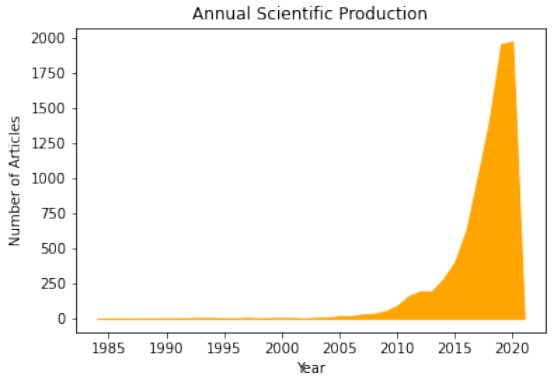

FIGURE 6: Annual Scientific Production

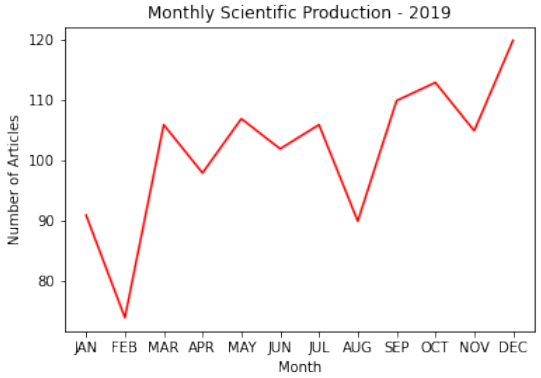

FIGURE 7: Monthly Scientific Production - Year 2019

that the start of the year 2020 was very productive but after June the graph shows the drastically decreasing trend. We can relate this trend to COVID-19 global pandemic affect. Total number of publication could be greater than the total achieved if there would be no pandemic.

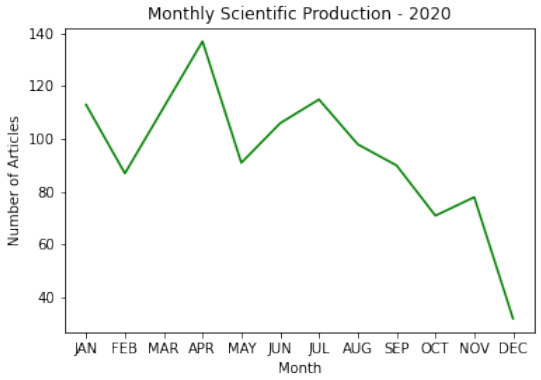

FIGURE 8: Monthly Scientific Production - Year 2020

D. AUTHORS CORRESPONDING COUNTRY

Web Of Information is additionally utilized to supply examination on International Collaboration by tallying number of articles published by creators inside the same country and papers with creators from other countries. In other words, there are two sorts of articles: single country publications (SCP) in which all creators belong to the same country and different country publications (MCP) in which creators belong to other countries.

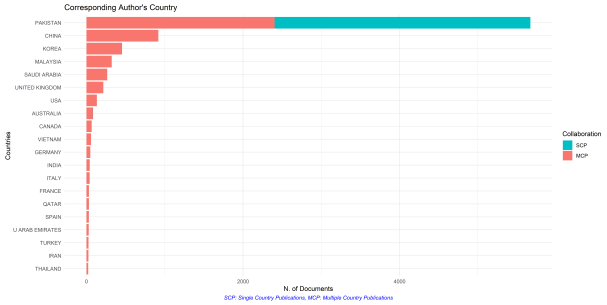

FIGURE 9: Authors Corresponding Countries

Figure No. 9 shows that the strongest literature production and collaboration on the field of Computer Science. China, Korea, Malaysia and Saudi Arabia have played an very important role in collaborating with Pakistan on literature production and scientific research. Figure No. 9 also shows that the most of the publications are written by authors belonging to the same country. This could be due to the fact that authors are keener to collaborate within the same research group or with academics with the same national background. We can also see that, Pakistani researchers are collaborating in very few projects with Neighbour countries like India, Iran and Afghanistan.

From Figure 20, We can see that the top funding organizations for research in Pakistan are Chinese, Korean and Saudi Arabian. We can conclude from that this high collaboration with these countries is due to the high funded projects in Pakistan. There are also multiple scholarships programs between these countries and very large scale of Pakistani researcher pursuing their PhD's from these countries. These types of programs and collaborations are helping Pakistani researcher to grow and get up-to-date in Research Industry. These findings could help us to find most appropriate collaboration countries.

E. NETWORK GRAPH ANALYSIS

The extraction of semantic relation plays an important role. The semantic relation between keywords helps us to establish facts about the connections between the keywords. These analyses also help researchers to find domains and sub domain that have a strong relation to each other. The two words said to be co-occurred if they appear in same article. By linking co-occurrences, a co-occurrence graph is formed and co-occurrence pair are also called neighbors. Terms in larger font represent the higher co-occurrence than other terms. We can see in Figure No. 10 below there are three main clusters red, blue and green. Distance between nodes provide relationship between them. The biggest node is machine

learning in red color cluster and its closest neighbors are deep learning, feature extractions and classification. In second cluster (blue) biggest node is cloud computing and its closest neighbor are internet of things, big data and security. In third cluster (green) wireless sensor network, energy efficiency and routing are neighbors.

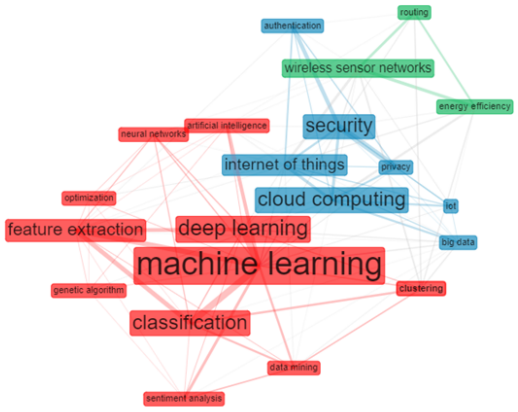

FIGURE 10: Author Keywords Network

F. KEYWORD GROWTH

Abstract is the rundown of the parts of the paper and is a time-saving easy route for active analysts to understand the paper. It may be a direct to the foremost imperative parts of your paper's composed content. Many people just read the Abstract section of paper to get idea of it. The abstract portion of the paper answers following questions :

- What was done in the paper?
- Why did we did what we did?
- What are your findings?
- Why are these findings useful and important?

Keywords are apparatuses making a difference indexers and look engines finding pertinent papers. On the off chance that databases and search engines are able to discover your paper, so users will be able to discover it as well. This will increment the number of individuals pursuing your paper, and likely lead to more citations.

- Keywords represents the content of the paper.
- Keywords are specific to the field or sub-field.

Abstract Word Growth shows the occurrences of words in the Abstract section of the papers in our dataset. The Abstract Word Growth, shown in Figure 11, demonstrates the growth of the words in the abstract section over the years. As evident from the figure, Proposed, Data, Paper are the most popular abstract words with more than 1000 occurrences annually. However, according to the figure, these abstract words obtained popularity after 2008. According to the forecast, these abstract words will see increase in growth as well in the coming year.

Author Keywords are a significant words, or a words used to find information when researching a specific paper.

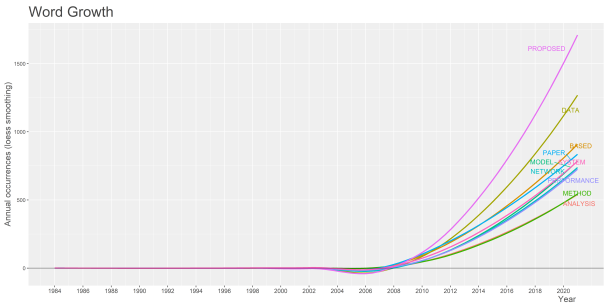

FIGURE 11: Abstract Word Growth

Keywords represent the domain of the paper and they are specific to field or sub-field of study of the paper. Author keywords are represented by a KEYWORDS in Research Publication after the Abstract of the publication. The main purpose of keywords is to help viewers to find the paper in the relevant research publications list. Keywords helps in indexing of the paper and it makes the paper more search able. That's why it is important for Authors to mention the most relevant keywords that will help other authors find your paper.

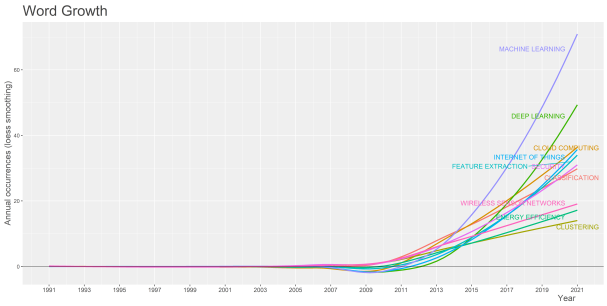

FIGURE 12: Author Keyword's Word Growth

Author Keyword Growth shows the occurrences of keywords in the Author Keyword section of the papers present in our dataset. The Author Keyword Growth, shown in Figure No. 12, demonstrates the growth of the Keywords in Author Keyword section over the years. It also shows which topic was researched the most over the period because Keywords gives us the idea of the paper so from the figure we can conclude on which topics and fields the research took place. As evident from the figure, MACHINE LEARNING is the most popular Author Keywords with more than 60 occurrences annually. Following MACHINE LEARNING is DEEP LEARNING with almost 50 occurrences annually. From this observation, we can conclude that more research has been focused towards Data Science. However, according to the figure, these abstract words obtained popularity after 2012. According to the forecast, these abstract words will see increase in growth as well in the coming year. Also there is another observation which shows that study of papers on topics relating CLUSTERING and ENERGY EFFICIENCY has slowed down over the years. Author Keywords has most

specific words that maps the paper with that keyword

The title of the paper is usually the first introduction readers (and reviewers) have of the paper. Therefore, a title is that grabs attention, title accurately describes the contents of the paper, and makes people want to read further. Title conveys the main topics of the paper and highlights the importance of the research paper. It is concise and it attract readers. The title lists the topics covered by the research paper. Authors tries to put all of the topics together in the title using as few words as possible. Title of a paper creates interest in other research scientists to review and study your paper. This should be related to the reason you decided to study the topic. Finding out and searching famous subject is thought as one of the amazing ways to study paradigm changes in any research field because of this we can know the latest research trends of a field.

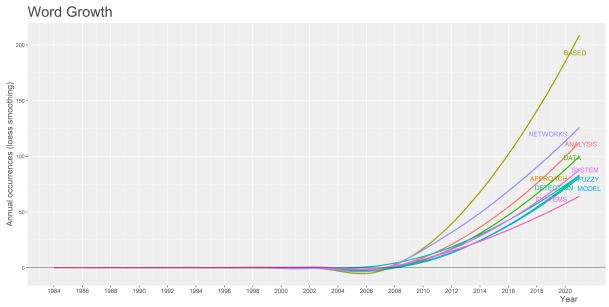

FIGURE 13: Title Word's Growth

Title Keyword Growth shows the occurrences of keywords in the Document Title of the papers present in our dataset. The Title Word Growth, shown in Figure No. 13, demonstrates the growth of the words of Document Title over the years. As evident from the Figure No. 13, BASED is the most popular Title Words with more than 200 occurrences annually. Following BASED is NETWORK with almost 125 occurrences annually. From this observation, we can conclude that most research document titles have based or network or analysis related word in Document Title. However, according to the figure, these words obtained popularity after 2008. According to the forecast, these abstract words will see increase in growth as well in the coming year.

G. WORD CLOUD ANALYSIS

For the better visualization of Bibliometric Analysis three word clouds are presented in this section. The purpose behind the use of word cloud for papers in field of computer science is that to analyze the most predominant words which represents most of the work is in those areas. The Figure No. 14 below shows the trending topics in the field of computer of science from year 1984 to 2021. Deep learning, cloud computing, classification and feature extraction are the most relevant topics. Security, image processing, big data and clustering are also in the list.

The Figure No. 15 below is the word cloud of titles most prominent words in the field of computer science from year

FIGURE 14: Author Keywords Word Cloud

1984 to 2021. Networks, data, based and analysis are most relevant words in the title word cloud.

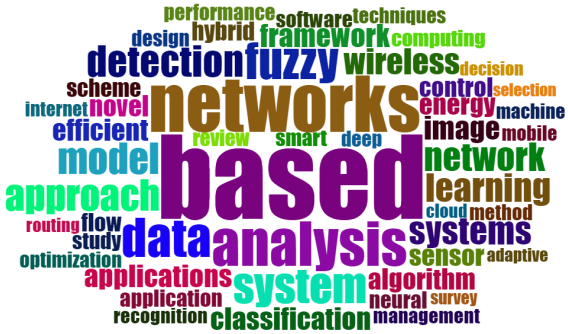

FIGURE 15: Title Word Cloud

Abstract of a paper contains the summary of paper and it contains the important information of your paper. Below Figure No. 16 is the word cloud of abstract. Proposed, network, systems, techniques, model and algorithm are the prominent words.

FIGURE 16: Abstract Word Cloud

H. TOP JOURNALS

A journal is a scholarly publication containing articles written by researchers, professors and other experts. Journals focus on a specific discipline or field of study. Unlike newspapers and magazines, journals are intended for an academic or technical audience, not general readers.[14] From Figure No. 17, we can see that the The Institute of Electrical

and Electronics Engineers Access (IEEE Access) is the top journal with 1400+ publications. International Journal of Advanced Computer Science & Applications and International Journal of Computer Science & Network Security are also the promising journals where majority of the documents published by Pakistani authors. International Journal of Distributed Sensors Networks and Computer and Electrical Engineering Journal are the most declined journal. This also shows the trend of research that Pakistani researchers research in domain of Sensor Network and Electrical Engineering is very low.

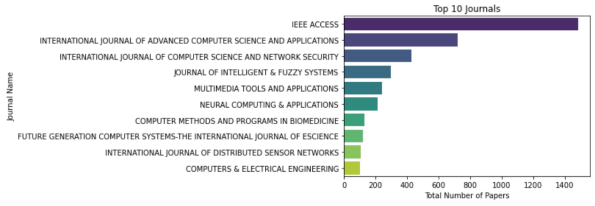

FIGURE 17: Top 10 Journals

The Source Growth, shown in Figure 18, demonstrates the growth of the journal sources over the period. As evident from the figure, IEEE Access is the most popular journal with annual occurrences of about 350 occurrences per year. However, according to the graph, the source obtained popularity after 2010, while before that, International Journal of Advanced Computer Science and Applications and International Journal of Computer Science and Network security, were the only most popular sources with annual occurrences of about 100 and 90 occurrences per year respectively. The source with the lowest growth potential shown up to the year 2020 is the Multimedia Tools and Applications with publication about 40. Furthermore, the journal International Journal of Advanced Computer Science and Applications is also not showing a very significant improvement and all others are also showing the same trend. IEEE Access is leading Journal and will be leading in future years.

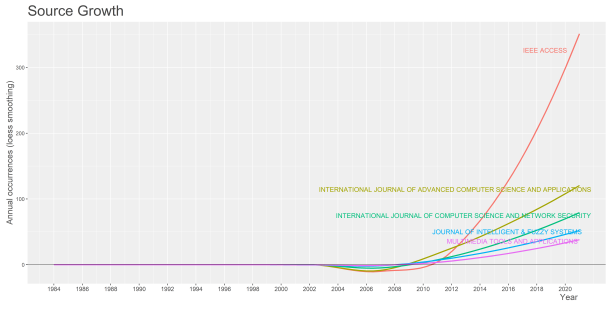

FIGURE 18: Title Word's Growth

I. TOP PUBLISHERS

Publishers are the organization which published the research papers. Publishers have different journals for specific type of research in which they publish research. From Figure No. 19 we can see that the IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC is the top publisher with around 1750 publications. SPRINGER is also the leading publisher with total publications of around 1000. Whereas ELSEVIER SCIENCE BV and WILY are the most declined publishers.

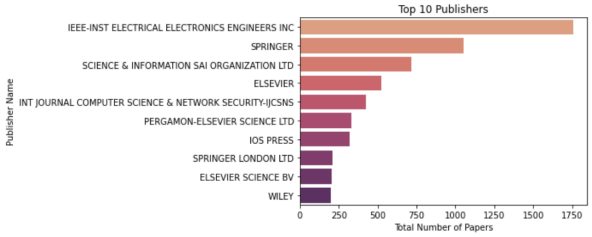

FIGURE 19: Top Publishers

J. TOP FUNDING ORGANIZATIONS

The important factor to impact the quality of research is Funding. Mostly the funded projects are the most focused and impactful projects. Funding Agencies normally fund the research groups for high end research on some research topic. From Figure No. 20, We can see that the National Natural Science Foundation of China(NNSFOC) has funded the majority of the funded researches. This contribution by Chinese organization makes china the Top Corresponding Author country we can observe that in figure. Higher Education Commission (HEC) has funded around 480 research projects and leading the table. National Research Foundation of Korea and King Saud University of Saudi Arabia are the top funders in Pakistani research which also make these country the Top Corresponding Authors Countries.

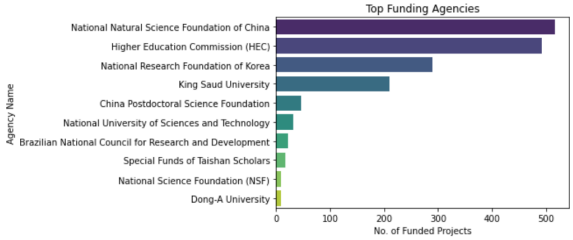

FIGURE 20: Top Funding Agencis

V. CONCLUSION

This paper shows a Bibliometric Analysis of the research work done by Pakistani researchers in the domain of Computer Science over the period from 1984 to 2021. We found out that the top funding agencies in terms of funding projects related to research in the field of Computer Science were the National Natural Science Foundation Of China with 500 funded projects followed by HEC having greater than 450 funded projects and at third and fourth comes National Research Foundation Of Korea with 300 funded projects and then King Saud University, Saudi Arabia with almost 250 funded projects. We also did find Single Country Publications

(SCP) and Multi-Country Publications (MCP) in Author Corresponding Country.

We found out the Annual Scientific Production from the number of articles over the years and we also found out the Citation Distribution from the number of articles. Further, Author Keyword, Title Word Growth, Abstract Word Growth, Author Keyword co-occurrence Network is also discovered. Bradford's Law is used to determine the number of core journals and articles in the Computer Science domain. Lotka's Law is used to describe the frequency of publication by authors in the Computer Science domain.

From most frequent Author Keywords, The topic's trending greatly affects the number of publications and citations in that field. We notice that topics associated with Data science, such as Machine Learning, Deep Learning, and Cloud Computing, have been trending. We also believe that interest in Data Science will keep on growing in the comings years.

We also found out Top Authors by their publications over the years. KHAN. M is the top author with more than 200 articles publications in 2020. So the most productive author in terms of publications is KHAN.M. Citation analysis is used to measure the impact of papers, indicating field usage and value carried by that article. This can be a very useful measure to assess the yearly impact of articles. Citation Analysis tells us the total number of citations present against a Keyword. Network Graph Analysis was used for the extraction of semantic relation. The co-occurrence network graph is used to extract the semantic relation between the author keywords.

Word Cloud Analysis is used for better visualization of Bibliometric Analysis content. The purpose behind the use of word cloud in this paper was to analyze the most predominant words which represent most of the work is done in which fields. Meta-Based was used to determine the number Of publications per author and count Of publications per country and n-index of author tells that n papers of researcher got n citations and count Of References per article For Content-Based Analysis, Word Cloud was used to determine the most frequent word used in Document Abstract, Author Keywords, and Document Title.

REFERENCES

[1] Iqbal, W., Qadir, J., Tyson, G. et al. A bibliometric analysis of publications in computer networking research. Scientometrics 119, 1121–1155 (2019).

[2] Ali, Muhammad & Richardson, Joanna. (2016). Research Publishing by Library and Information Science Scholars in Pakistan: A Bibliometric Analysis. Journal of Information Science Theory and Practice. 4. 6-20. 10.1633/JISTaP.2016.4.1.1.

[3] Bajwa, R.S., Yaldram, K. Bibliometric analysis of biotechnology research in Pakistan. Scientometrics 95, 529–540 (2013). https://doi.org/10.1007/s11192-012-0839-x.

[4] V. Garousi and J. M. Fernandes, "Highly-cited papers in software engineering: The top100," Information and Software Technology, vol. 71, pp. 108-128, 2016.

[5] Muhammad Kamran, Hikmat Ullah Khan, Wasif Nisar, Muhammad Farooq, Saeed-Ur Rehman, Blockchain and Internet of Things: A bibliometric study, Computers & Electrical Engineering, Volume 81, 2020, 106525, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2019.106525.

[6] Missen, M.M.S., Qureshi, S., Salamat, N. 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, 113–142 (2020). https://doi.org/10.1007/s11192-020-03379-8.

[7] Fiala, Dalibor & Tutoky, Gabriel. (2017). Computer Science Papers in Web of Science: A Bibliometric Analysis. Publications. 5. 10.3390/publications5040023.

[8] Bakri, A., & Willett, P. (2011). Computer science research in Malaysia: a bibliometric analysis. Aslib Proceedings, 63(2/3), 321–335. doi:10.1108/00012531111135727.

[9] Y. Ding, G. Zhang, T. Chambers, M. Song, X. Wang, and C. Zhai, "Content-based citation analysis: The next generation of citation analysis," Journal of the Association for Information Science and Technology, vol. 65, pp. 1820-1833, 2014.

[10] J. E. Dayhoff, Neural network architectures: an introduction: Van Nostrand Reinhold New York, 1990.

[11] www.webofknowledge.com

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

[13] https://clarivate.com/webofsciencegroup/solutions/web-of-science

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