Journal of Scientometric Res. 2022; 11(1):37-46
http://www.jscires.org

Research Article

Stress and Machine Learning-Future with Possibilities: A Bibliometric Approach

Pooja Gupta[1,*], Srabanti Maji[1], Ritika Mehra[2]
[1]*School of Computing, DIT University, Dehradun, Uttarakhand, INDIA.*
[2]*Devbhoomi Uttarakhand University, Uttarakhand, INDIA.*

ABSTRACT
Stress in human life is a global health concern and machine learning based models have been applied extensively for the stress prediction. This work is an attempt to present a bibliometric analysis in the field of stress prediction using machine learning. The dataset to conduct this study was taken from the Web of Science database and research papers were selected from the year 2005 to 2021. Then, bibliometric analysis tool, VOS Viewer 1.6.14 was applied for generating a co-authorship network map, inter-country co-authorship network map, and keywords co-occurrences network maps. The outcomes of this study visually highlight the important research details like the most prolific journal, most cited paper, most prolific country, institution and interesting research driving points in the stress prediction using machine learning. This study attempts to portray the existing literature on stress prediction using machine learning more comprehensively and systematically by showing research collaboration among countries, authors, co-citations analysis and, bibliographic coupling. The findings of this study can be useful to conduct future research on a similar area.
Keywords: Bibliometric analysis, Stress prediction, Machine learning, VOSviewer, Bibliographic coupling.

Correspondence
Pooja Gupta
School of Computing, DIT University, Dehradun, Uttarakhand, INDIA.
Email id: pooja.gupta@dituniversity.edu.in

Received: 14-09-2021;
Revised: 15-11-2021;
Accepted: 16-01-2022;
DOI: 10.5530/jscires.11.1.4

INTRODUCTION

The most complex part of the human body is its brain with nearly 100 billion neurons, it is an interpreter of senses, responsible for body movement, and controller of human behaviour. It is also a key organ of the response to stressors. During its evolution, a series of genetic, work-related, family, and other social-environmental stressors can cause structural and functional abnormalities.[1] Stress is defined as the body's nonspecific response to any demand for change.[2] Prolonged stress can result in an abrupt reduction in brain weight and brain mass.[3] Therefore, the evolution of various strategies for early prediction of stress plays crucial role in health care sector of late, machine learning based models have been used extensively for stress prediction and various strategies for stress prediction have been reported in the literature such as questionnaire-based technique where, the psychiatrist gives the user a feedback form, and then, depending on the information gathered, determines whether or not the user is stressed. Sensor measuring technique which is based on based on galvanic skin response and skin temperature and machine learning algorithms. Stress discovery using social media data where social media data can be analysed to learn more about issues, trends, key actors, and other topics. Stress analysis using facial feature recognition which detects stress levels from particular facial cues that are generated from eye, mouth, and head motions. A study presented a review of different stress assessment techniques and concluded that techniques which includes SVM, decision trees, and Bayesian networks have been applied by most of the researchers.[4] Another study on various Facial Emotion Recognition system was presented by,[5] where authors focussed on automatic facial recognition and presented a brief overview of various process, techniques, and application of facial emotion recognition system. Another study[6] stated that people on a daily basis, are exposed to mental stress due to a variety of factors such as traffic, noise, and bad weather, as well as social aspects, are all factors to consider. Authors provided a review on stress detection where focus was on facial features extraction, wearable sensor devices and biosignal processing. In another review article, authors[7] investigated the stress detection methods with sensory equipment such wearable sensors, electroencephalography (EEG), electrocardiogram (ECG), and photoplethysmography (PPG), as well as different contexts including driving, learning, and working. All these mentioned studies lack a comprehensive review of recent research inclinations and patterns of this field. Hence, the focus of this research is to conduct a bibliometric analysis from a different perspective by mapping the research in this area using the software VoSViewer.

Bibliometrics is the interdisciplinary discipline of quantitative analysis of all the available literature performed by mathematical and statistical tools.[8,9] Considering bibliometric strategies, the most recent advances, driving points, and research gaps in a specific research field can be drawn, and it is turning into an important research technique for surveying national and global research efficiency, worldwide collaboration, research patterns, citation analysis, and improvement of particular areas. Lately, numerous bibliometric investigation approaches and tools such as CiteSpace and VOSviewer have been advanced to support researchers in various fields to generate information maps, assess the overall state of thought about a specific research topic, and recognize hotspots in the research area.[10] Bibliometric analysis on mental health has been done by many authors.[10-13] An extensive review of all the accessible literature suggested that previous studies conducted only some very preliminary descriptions. Previous bibliometric analyses were only descriptive and straightforward, focusing on the most cited publications/journals, the countries of most significant collaboration. Proposed study discloses new insights and is different from other works in the same field as it presents a complete bibliometric analysis in order to objectively evaluate the new insightful maps and contribution of global researchers in the field of stress prediction using machine learning by providing following findings:

1. Journals with the highest number of publications
2. Most cited research papers
3. Prolific countries and organizations
4. Authors who have co-operated the maximum, with other authors
5. Countries whose authors have co-authored the maximum with the authors of other countries
6. Co-citations: Cited Authors
7. Co-citations: Source
8. Most recurrently occurring keywords in the research field
9. Bibliographic Coupling: Document-wise

Outline shows the organisation of this paper: Segment 2 describes the research methodology and flow of this work, the results, and findings are presented in segment 3. Finally, the conclusion and discussion are provided in the last segment 4.

RESEARCH METHODOLOGY

The main determination behind conducting this research is to conduct a bibliometric analysis to analyse the work done on stress prediction using machine learning, from the year 2005 to 2021. Bibliometric analysis is a quantitative analysis that is conducted to find out the description of several features of published literatures in a particular research area and to provide a simple way to examine the current status and forecast the developmental trends in a specific research area.[14-16] A quantitative analysis of a research discipline is yet another advantage of bibliometric analysis.[17] For searching and selection of research papers related to stress prediction using machine learning the Web of Science (WoS) database was used and the time-period of paper publication was set to the year 2005-2021. The reason behind selecting the WoS database for analysis is simple as WoS is utilized as both a dataset for data-oriented research and as a tool to conduct academic library research. The complete research methodology is explained in Figure 1.

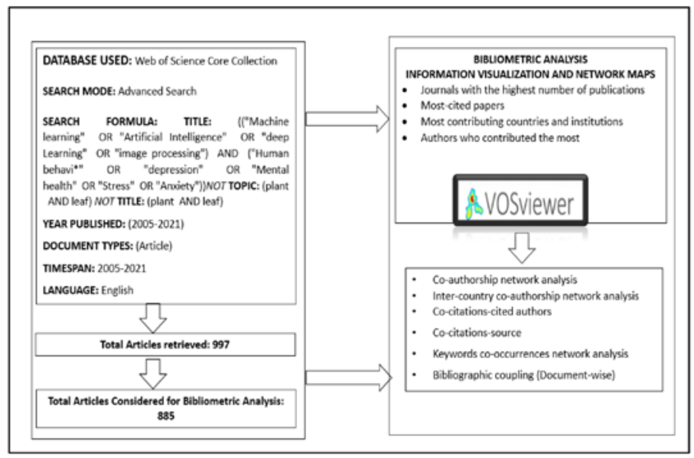

Figure 1: Flow of Research Methodology.

WoS contains a huge number of bibliographic annals involving billions of reference associations and extra metadata fields; and a large number of added things are ingested every day.[18] To select the research papers on the area of "stress prediction using machine learning" a search filter was applied and the filter was moderated to select only those articles which included the keywords "Machine Learning" or "Artificial Intelligence" or "Deep Learning" or "image processing" and "mental health" or "stress" or "depression" or "anxiety" or "human behave*" in either their title or topic from the year 2005 to 2021. Further, the search produced a total of 997 articles. Among those publications, the three main document types were: Articles (*n*=892, 89.46%), review (*n*=88, 8.82%) and other types like letters and editorial material were less than one percent. Authors read and scanned the title, abstract, and keywords of every research paper reflected in the search results to verify if papers were related to the topic or not. Consequently, out of the 997 papers, 105 papers were excluded as they belonged to slightly distinct topics. Further articles written in only English language are consider for this study, henceforth the remaining 885 papers were considered to conduct this bibliometric analysis. The bibliometric analysis tool VOS Viewer 1.6.14. was used to conduct co-authorship visualization analysis, country co-authorship network analysis, co-citations network, co-keyword network analysis and bibliographic coupling. VOS Viewer is a free tool applied to perform bibliometric maps in an easy-to-understand format.[19] To create networks of scientific publications and journals, VOSviewer, an open-source software was applied for generating and viewing bibliometric networks. Each visualisation map is made up of a network of entities, or objects of interest. Edges indicating citation (co-citation and bibliographic coupling), co-authorship, and co-occurrence linkages connect items such as research documents, authors, and keywords with other entities. Each link has a strength value connected with it, which is expressed as a positive number. The stronger the link between the connected entities, the higher this value.[8]

RESULTS AND FINDINGS

Journals with the Highest Number of Publications

Selected 885 papers were from 347 different journals. Figure 2 represents the top ten publishing journals with the maximum number of research papers on the topic of stress prediction using machine learning. The top journal publishing in this area is “IEEE Access” with 51 publications which is 5.717% of total publication. Next top journal publishing in this area is “Journal of Affective Disorders” with 43 publications which is 4.821% of total publication. This journal belongs to Elsevier publishers. Total four journals from the top ten belong to the Elsevier publishers, another two belong to the Frontier Media, and the rest three belong to other publishers like BioMed Central, Nature Publishing Group, American Medical Association.

Most Cited Papers

Table 1 presents the ten most cited papers in the specified field and the list of these papers is prepared from the dataset extracted from the WoS database. With the highest share of

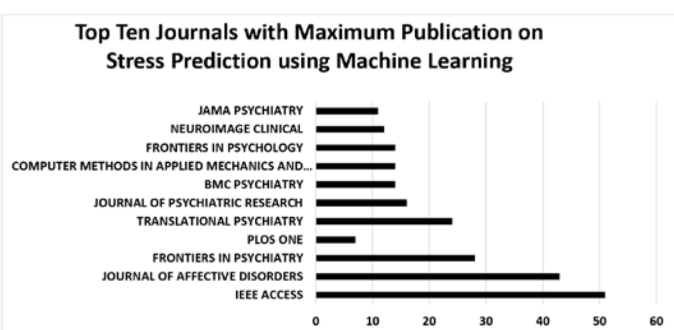

Figure 2: Top Ten Journals.

352 citations, the most cited paper is by Chekroud AM *et al.* (2016).[20] In this research paper, the authors established a model based on the 25 most crucial factors that came out as the main predictive of treatment response from 164 patient-reportable factors. Further, this model was validated, and anticipated results in the STAR*D cohort with precision above chance (64·6% [SD 3·2]; p<0·0001).

The second-most cited paper with 300 citations by Mohr *et al.* (2017)[21] represents a critical analysis of mental health issues through personal sensing research. This work critically reviews the use of smartphones, but articles on wearables, social media, and computers were also included. Further, with 258 citations, the third-most cited paper is by Hosseinifard *et al.* (2013).[22] This study shows that EEG nonlinear analysis is effective in discriminating between depressed patients and regular subjects. This research was conducted on 90 subjects, out of which 45 were normal subjects and 45 were unmedicated depressed patients. The authors concluded that the nonlinear analysis of EEG may be an effective tool to support the diagnosis of depressed patients. Furthermore, with 165 citations the research paper by Nouretdinov et al. (2007) [23] is at the fourth position, this paper landscapes a general technique of probabilistic arrangement to generate measures of confidence for magnetic resonance imaging data. Further, with 66 citations the fifth-most cited paper is Koutsouleris *et al.* (2018) with 159 citations.[24]

Sixth most cited paper with 118 citations is by Mansson *et al.* (2015),[25] this paper focuses on participants suffering from social anxiety disorder after completing one year of Internet-delivered (CBT) i.e., iCBT, and evaluate neural predictors of long-term treatment in contributors.

Further, the three most cited papers are by Helbich *et al.* (2019), Patel *et al.* (2016, 2015) with 118,102 and 90 citations respectively.[26-28] Helbich *et al.* (2019)[26] studied mental and emotional disorders affecting older people i.e., geriatric depression in Beijing, China. Patel *et al.* (2016)[27] presented a review of past work done on mental health, depression, imaging, and machine learning methodologies. Further, the next four most-cited research articles in this are Lin *et al.* (2018), Galatzer *et al.* (2018), and Oliveira *et al.* (2013).[29-31] Lin *et al.* (2018)[29] proposed a model based on deep learning algorithms that differentiate responders and non-responders, and the model is also capable of predicting antidepressant treatment outcomes in major depressive disorder (MDD). Galatzer *et al.* (2018)[30] examined the possible outcome of novel analytic tools and then further re-examined the data gathered from the peoples who were diagnosed with trauma that was unsuccessful to spot a linear relationship between traumatic events and cortisol response and afterward post-traumatic stress disorder. Oliveira *et al.* (2013)[31] studied the pattern of brain activation and suggested that there is a

Table 1: Ten Most Cited Paper.

#	Title	Authors	PY*	TC*	Avg/year
1	Cross trial prediction of treatment outcome in depression: a machine learning approach	Chekroud, Adam Mourad; Zotti, Ryan Joseph; Zarrar; Shehzad; Gueorguieva, Ralitza; Johnson, Marcia K.; Trivedi, Madhukar H.; Cannon, Tyrone D.; Krystal, John Harrison; Corlett, Philip Robert	2016	352	58.66
2	Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning	Mohr, David C.; Zhang, Mi; Schueller, Stephen M.	2017	300	60
3	Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal	Hosseinifard, Behshad; Moradi, Mohammad Hassan; Rostami, Reza	2013	258	28.66
4	Machine learning classification with confidence: Application of transudative conformal predictors to MRI-based diagnostic and prognostic markers in depression	Nouretdinov, Ilia; Costafreda, Sergi G.; Gammerman, Alexander; Chervonenkis, Alexey; Vovk, Vladimir; Vapnik, Vladimir; Fu, Cynthia H. Y.	2011	165	16.5
5	Prediction Models of Functional Outcomes for Individuals in the Clinical High-Risk State for Psychosis or With Recent-Onset Depression a Multimodal, Multisite Machine Learning Analysis	Koutsouleris, Nikolaos; Kambeitz-Ilankovic, Lana; Ruhrmann, Stephan; Rosen, Marlene; Ruef, Anne; Dwyer, Dominic B.; Paolini, Marco; Chisholm, Katharine; Kambeitz, Joseph; Haidl, Theresa; Schmidt, Andre; Gillam, John; Schultze-Lutter, Frauke; Falkai, Peter; Reiser, Maximilian; Riecher-Rossler, Anita; Raimo K. R.; Upthegrove, Rachel; Hietala, Jarmo; Salokangas, Pantelis, Christos; Meisenzahl, Eva; Wood, Stephen J.; Beque, Dirk; Brambilla, Paolo; Borgwardt, Stefan	2018	159	39.75
6	Predicting long-term outcome of Internet-delivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning	Mansson, K. N. T.; Frick, A.; Boraxbekk, C-J; Marquand, A. F.; Williams, S. C. R.; Carlbring, P.; Andersson, G.; Furmark, T.	2015	118	16.85
7	Using deep learning to examine street-view green and blue spaces and their associations with geriatric depression in Beijing, China	Helbich, Marco; Yao, Yao; Liu, Ye; Zhang, Jinbao; Liu, Penghua; Wang, Ruoyu	2019	118	39.33
8	Studying depression using imaging and machine learning methods	Patel, Meenal J.; Khalaf, Alexander; Aizenstein, Howard J.	2016	102	17
9	Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction	Patel, Meenal J.; Andreescu, Carmen; Price, Julie C.; Edelman, Kathryn L.; Reynolds, Charles F., III; Aizenstein, Howard J.	2015	90	12.85
10	A Deep Learning Approach for Predicting Antidepressant Response in Major Depression Using Clinical and Genetic Biomarkers	Lin, Eugene; Kuo, Po-Hsiu; Liu, Yu-li; Yu, Younger W. -Y.; Yang, Albert C.; Tsai, Shih-Jen	2018	66	16.5

*PY Publication Year, *TC Total Citations

difference in the brain activation pattern to normal expressions in mentally upset subjects with the pattern found in healthy control subjects to similar stimuli. Authors interpreted that this variation in brain activation may inspire the behavioural misinterpretation of the neutral faces content by the mentally upset patients.

Another major work in this area has been conducted by Lee *et al.* (2018), Gkotsis *et al.* (2017), and Michie *et al.* (2017).[32-34] Lee *et al.* (2018)[32] searched Ovid MEDLINE/PubMed database relevant studies and focussed on adults with bipolar or unipolar depression, further the authors applied a machine learning algorithm. Gkotsis *et al.* (2017)[33] applied 11 disorder themes to develop a classifier to recognize and categorize posts from the social media platform Reddit related to mental sickness. Michie *et al.* (2017)[34] worked on The Human Behaviour-Change Project (HBCP), HBCP developed a 'Knowledge-based System' using machine learning techniques. This project worked to automatically extracts and generate knowledge about behaviour change.

Most Contributing Countries and Institutions

Figure 3 represents ten topmost countries with the highest number of research articles related to stress prediction using machine learning. According to WoS, all the 885 research papers included are written by their respective authors of 74 diverse countries. In the case of multiple authors, the country of the first author was acknowledged as the country of source for the research article. Out of these 74 countries, the United States of America is at the first position with a maximum of 415 papers. China, England, Australia, and Germany are in second, third, fourth, and fifth positions respectively. One of the interesting observations is that out of 885 papers in the study; more than 46.52% of work is contributed by the authors of the United States only. Further to mention more than 76% of papers are contributed by the top five countries

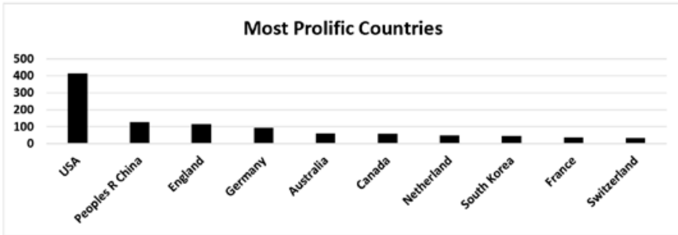

Figure 3: Most Prolific Countries with number of publications.

Figure 4: Most Prolific Organization.

i.e., the USA, China, England, Australia, and Germany. Figure 4 shows the most important institutes with maximum publications. The Kings College London, UK is the most important organization with 37 papers. Harvard Medical School is in second position with 33 publications. Columbia University is in the third spot with a total of 3 publications. University of Melbourne and University of Pennsylvania are at fourth position with 23 publications each.

Co-Authorship Visualization Analysis

Co-authorship analysis is the research collaboration among authors. Co-authorship analysis is performed on the research articles collected from WoS and a visual map is shown in Figure 5(a). VOSviewer 1.6.14 is applied to generate co-citations maps of authors or journals or to generate a co-occurrence network of keywords from the co-occurrence data.[19] VOSviewer implements a VOS mapping feature to create a map[35,36] have given a detailed mathematical explanation of VOS mapping techniques. The purpose of conducting co-authorship analysis is to present the authors co-operation pattern. This analysis includes a total number of 2495 authors. A co-authorship network map was created by setting up the type of analysis as co-authorship, and the unit of analysis used is authors. In this analysis, the maximum number of authors allowed per document was set to 20. The minimum threshold of documents of an author allowed was set to 3, then out of 2495 authors, all 299 met this criterion. Figure 5(a) shows circles connect with several lines. Bigger

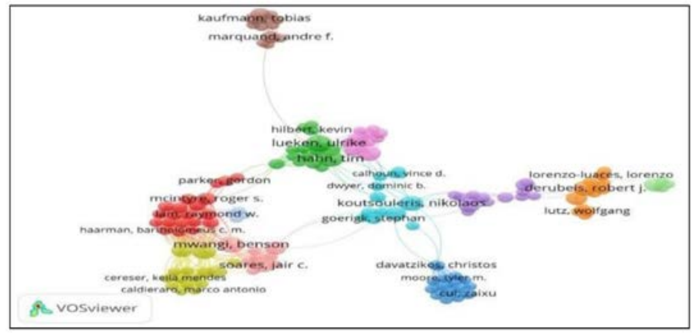

Figure 5: (a) Co-authorship network map of authors involved in the research area of Stress prediction using Machine learning.

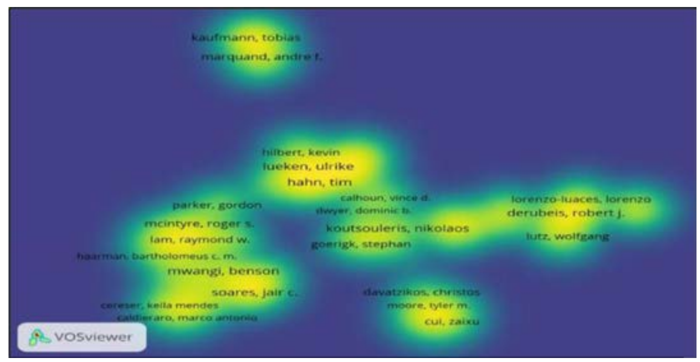

Figure 5: (b) Density visualization map based on co-authorship.

circles and labels are used to represent the name of authors who have co-operated maximum with the other authors.

Total 12 clusters are shown in Figure 5(a). With 20 authors the Cluster 1 in red colour, is identified as the biggest cluster and Cluster 2 represented by green colour, has 15 authors. There are 15 authors in cluster 2 in blue colour. Cluster 4 and 5 contains 15 and 14 authors respectively. The number of co-authorship relationships between a researcher and other researchers is shown by the links property. The Total link strength characteristic represents the total strength of a researcher's co-authorship ties with other researchers. Kapczinski, Flavio has collaborated with 35 distinct authors mood disorder, bipolar disorder, bipolar depression and machine learning, deep learning, unsupervised and semi -supervised machine learning. Figure 5(b) shows the density visualization map based on co-authorship. Table 2 represents strong co-authorship linked top 20 authors.

Country Co-Authorship Network Analysis

This analysis links different countries whose authors are co-authoring with each other. Country co-authorship network analysis provides country-wise co-operation of authors. All 885 research papers in the dataset originate from different 74 countries. Authors of 35 countries out of these 85 were associated through co-authorship. It has been further

Table 2: Strong Co-authorship linked top 10 authors.

#	Author	Document	Link	Total Link Strength
1	Kapczinski, flavio	6	35	78
2	Librenza-garcia, diego	5	32	65
3	Mwangi, benson	7	27	63
4	Kasper, siegfried	6	12	47
5	Kautzky, alexander	6	12	47
6	Souery, Daniel	6	12	45
7	Mendlewicz, julien	5	12	44
8	Montgomery, stuart	5	12	44
9	Serretti, alessandro	5	12	44
10	Zohar, joseph	5	12	44

observed that the leading countries in terms of co-authorship are the United States of America, Peoples R China, England, Australia, and Germany. It is visible from Figure 6, that the biggest circle is assigned to the "USA". Countries whose authors have cooperated a greater number of research articles are allocated thicker lines. Top 10 countries are represented in Table 3.

Co-citations: cited authors

If a publication cites two publications together then both these publications are said to be co-cited.[37,38] The stronger the co-citation relationship between two publications, the greater the number of publications by which they are co-cited. In the visualization, the distance between two writers roughly shows their relatedness in terms of co-citation relationships. Generally, the closer two authors are to one another, the more related they are.

Figure 7(a) shows 9 clusters, cluster 1 and 2 are the biggest clusters with red and green colours with 19 authors each

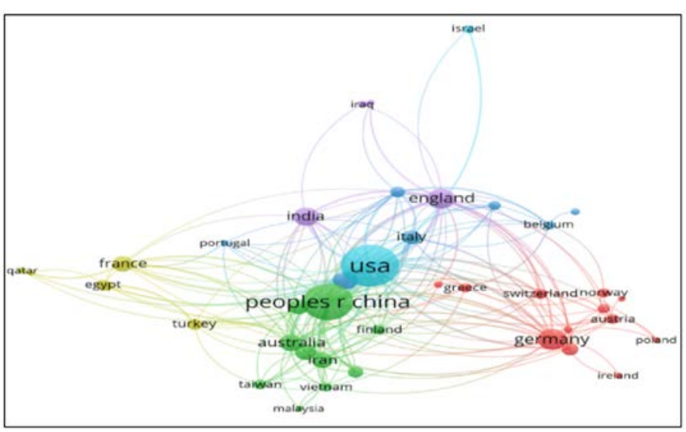

Figure 6: Country Co-Authorship Network Map.

Table 3: List of countries.

#	Country	N.C.
1	USA	141
2	Peoples R China	84
3	England	58
4	Australia	46
5	Canada	36
6	Germany	32
7	Iran	29
8	Japan	29
9	India	26
10	France	25
11	Netherlands	25
12	Italy	24
13	Spain	20
14	South Korea	17
15	Switzerland	17
16	Finland	15
17	Singapore	15
18	Turkey	15
19	Austria	14
20	Belgium	13

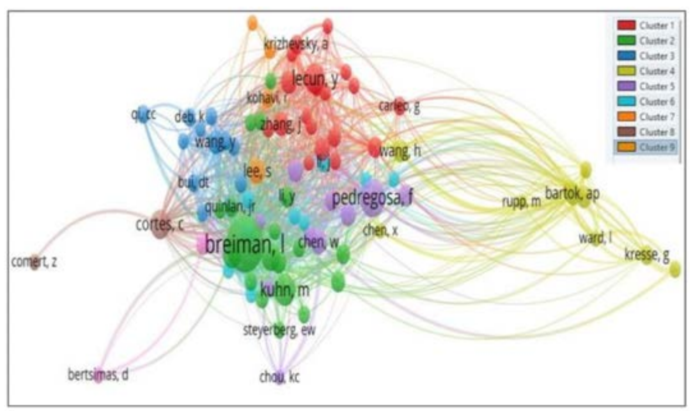

Figure 7: (a) Co-citations Visualization based on cited authors.

appear as the main fields, contributing to the stress prediction using machine learning literature. Cluster 1 in red colour consists of 19 authors and Cluster 2 in green colour also has 19 authors. Out of 19249 authors 34 met the criteria with minimum number of citations of an author=15. With total 127 citations and total link strength of 542 L. Breiman[39] has been co-cited the most. Out of the author co-citation network, it is also possible to observe Breiman is well connected with all nine clusters. Breiman proposed Random Forest in year 2006. This also shows that a lot authors have used Random Forest technique in stress prediction. Figure 7(b) shows the density visualization of co-citations based on cited authors.

Co-citation-source

In the graphic, the distance between two sources roughly shows their relatedness as per co-citation relationships. Generally, the closer two sources are to one another, the more related they are. Lines also show the strongest co-citation relationships between sources. Figure 8(a) shows total 5 clusters, cluster 1 is the biggest clusters in Red with 18 items. Second biggest cluster is in green colour with 11 items. Journal of Affective Disorder connected with all five clusters.

Co-Keyword Network Analysis

Co-keyword analysis also gives an idea about areas, methods, and topics that are often researched by the authors in the research. The main purpose of conducting the co-keyword network analysis is to Figure 8(b) out what are the keywords which have been regularly mentioned in several research articles. The text-mining technique is implemented by the software to generate a network map. In the visual map, the distance between different words is represented as the affinity among different keywords.[36] It is visible from the map that more distance between two or more keywords means these keywords are less related to each other and the lesser the distance, the stronger the term are related. Keywords occurring more than 5 times were selected from the dataset, 99 out of a total of 3364 keywords fulfilled this criterion.

The keywords "Machine learning", "Machine learning (ml)" and "Machine-learning" appeared a total of 249 times in the dataset. Other similar keywords to machine learning were also found in the dataset such as "Artificial Intelligence", "Deep Learning", "Neural-networks", "Artificial neural network" etc., in an observation if all the similar keywords are counted then a total of 366 keywords related to the keyword machine learning occurred in the dataset.

Keywords appearing a maximum number of times in the dataset, advocate the idea that most of the research has been conducted on these topics. If two keywords are connected through a line that means that these keywords have co-occurred with each other in several research articles in the dataset obtained from WoS. It has been noticed further from the Figure 9, that almost every other maximum appearing keyword in Table 4, has co-occurred with the term "machine learning" or with its synonymous terms. Figure 9 highlights

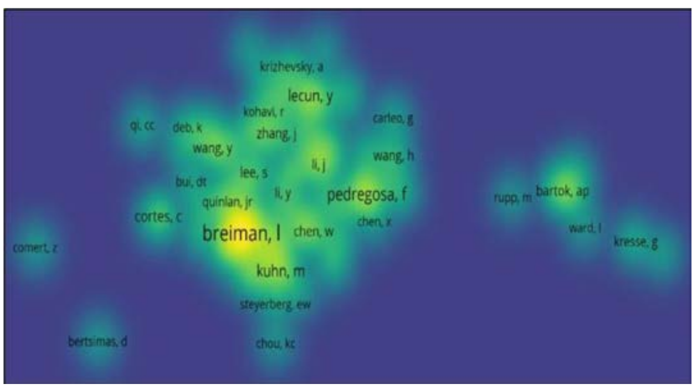

Figure 7: (b) Density visualization of co-citations based on cited authors.

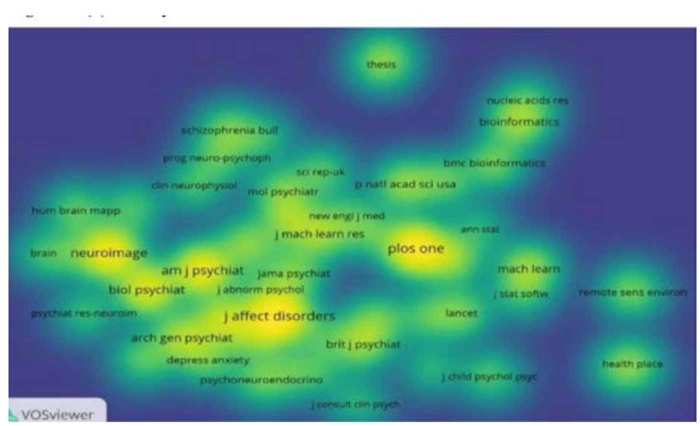

Figure 8: (b) Density visualization of co-citations based on cited sources.

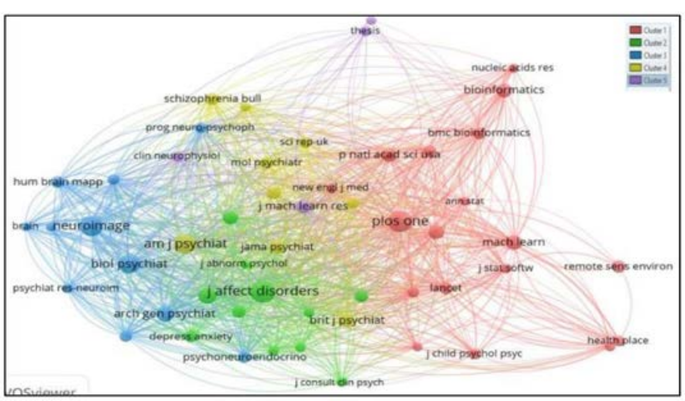

Figure 8: (a) Co-citations Visualization based on cited sources.

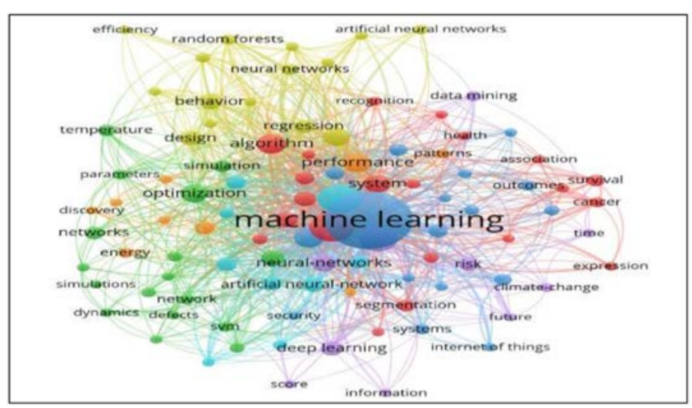

Figure 9: Keyword Co-Occurrences Network Map.

Table 4: Maximum appearing keywords in the dataset.

#	Keywords	N.C.
1	Machine learning	237
2	Prediction	60
3	Model	47
4	Classification	62
5	Random forest	29
6	Optimization	23
7	Algorithm	27
8	Support Vector Machine	24
9	Performance	25
10	Neural-networks	20
11	System	22
12	Artificial intelligence	18
13	Behaviour	18
14	Identification	17
15	Selection	15
16	Deep learning	16
17	Design	16
18	Regression	14
19	Artificial neural network	14
20	Artificial neural-network	14
21	Big data	13
22	Models	13
23	Neural-network	13
24	Diagnosis	12
25	Features	12
26	Networks	12
27	Risk	12
28	Validation	11
29	Impact	11
30	Genetic algorithm	10
31	Management	10
32	Segmentation	10
33	Simulation	10
34	Survival	10
35	Temperature	10

different keywords, which co-occurred with the keyword "machine learning".

A co-occurrence map can be useful to provide information like terms, techniques, or areas, which have been used by the researchers with the term machine learning. For example, the co-occurrence of the keyword "prediction" and "machine learning", suggests that many researchers have developed some prediction models using machine learning. Further, the co-occurrence of the keyword "classification" and "machine learning", suggests that the classification algorithms have been implemented by different researchers for formulating models related to machine learning.

Bibliographic Coupling (Document-wise)

Bibliographic coupling opposite of co-citation If a third publication is cited by both publications, they are considered bibliographically connected.[40] A coupling unit between two documents is a point of reference shared by both of them. Two documents are said to be bibliographically connected if both cite a common research document. In other words, the greater the coupling relationship between two papers, the more references they have in common. Document-wise bibliographic coupling visualizations are shown in Figure 10(a), (b) and (c). Chekroud (2016)[20] has emerged as strong document in the dataset.

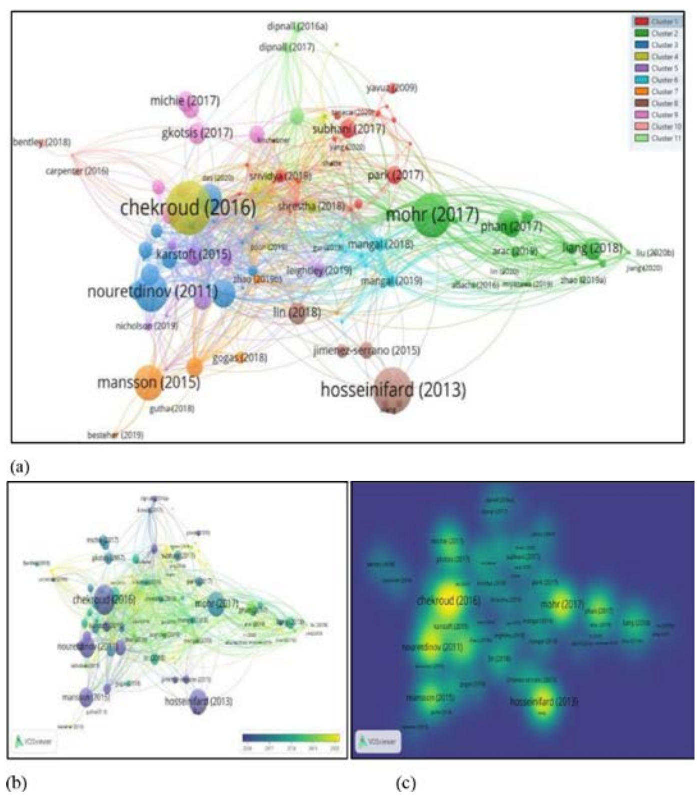

Figure 10: (a) Visualization of Bibliographic Coupling (Document-wise); (b) Overlay visualization of bibliographic coupling; (c) density visualization map of documents based on document weights.

DISCUSSION AND CONCLUSION

This bibliometric analysis shows that a lot of research has been conducted on stress prediction using machine learning from 2005-2021. This study successfully provides the answers to all research questions mentioned in section 1. Based on this bibliometric analysis the most contributing journal with the maximum number of articles on stress prediction using machine learning is "IEEE ACCESS", belonging to the "Elsevier publisher". This analysis provides many insights like the leading country in terms of maximum publications in the United States of America. Authors of the USA have also co-operated the maximum times, with the authors from other countries. The leading institute with the maximum number of publications under its name is King's College London, UK.

Based on this analysis there are 16 most prolific authors with three papers each in this area. Micheal Berk; Joanna Dipnall and, Denny Mayer are the top three authors. One more important observation was that out of 16 authors 7 prolific authors have co-authored the maximum times, with the other authors in the dataset obtained from WoS.

From author co-citation network, it is also possible to observe Breiman is well connected with all authors. Breiman proposed Random Forest in year 2006. This also shows that a lot authors have used Random Forest technique in stress prediction. In addition to this, another observation was made from keyword co-occurrence network that "Random Forest" is among the top 5 most occurring keywords in the dataset. Further, maximum occurring keywords in the dataset are "machine learning", "prediction", "model", "classification" and "random forest", as disclosed by the keyword co-occurrences network analysis. From bibliographic coupling and most cited paper analysis, it can be concluded that Chekroud (2016)[20] has emerged as strong document with 352 citations.

Apart from providing several critical information, this analysis has some limitations also, which provides a direction for future research. The main limitations are in terms of time-span chosen, database, language, type of publications. Talking about time-span limitations, papers published from January 2005 to May 2021 have been included in this analysis. In a future study, the time can be extended. The second limitation is that only Web of Science dataset was chosen for conducting this study, papers available in other databases like Scopus were not considered, a future study may use a dataset from Scopus also. The third limitation is that papers written in the English language are included only, hence a lot of papers written in other languages may have skipped. The fourth limitation is that while setting the filter for searching papers, authors have searched only for articles as document type, hence other important research published in form of doctoral dissertations, book publications were not included in this study. Henceforth, in future work, other document types may also be included and the obtained results can be compared with the results of this analysis.

With all these limitations this analysis is not complete in all parameters but it still manages to provide many useful insights. It can be concluded that this bibliometric analysis presents the literature available on stress prediction using machine learning all in one place. It is assumed that this study will be useful to the researcher who wants to study the research related to stress prediction.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Liang X, Wang JH. Human connectome: Structural and functional brain networks. Chinese Science Bulletin. 2010;55(16):1565-83. doi: 10.1360/972009-2150.
2. Selye H. The stress concepts. Canadian Medical Association Journal. 1976;115(8):718. PMID 20312787.
3. Sarahian N, Sahraei H, Zardooz H, Alibeik H, Sadeghi B. Effect of memantine administration within the nucleus accumbens on changes in weight and volume of the brain and adrenal gland during chronic stress in female mice, Pathobiology Research. 2014;17:71-8.
4. Deshmukh RS, Jagtap V, Paygude S. Facial emotion recognition system through machine learning approach. In: International Conference on Intelligent Computing and Control Systems (ICICCS). Vol. 2017. IEEE Publications; 2017;June:272-7.
5. Elzeiny S, Qaraqe M. Machine learning approaches to automatic stress detection: A review. In: 15th International Conference on Computer Systems and Applications (AICCSA). Vol. 2018. IEEE Publications/ACS. IEEE Publications; 2018;October:1-6.
6. Gedam S, Paul S. A review on mental stress detection using wearable sensors and machine learning techniques. IEEE Access. 2021;9:84045-66. doi: 10.1109/ACCESS.2021.3085502.
7. Johnson MH, Cohen J, Grudzinskas G. The uses and abuses of bibliometrics. Reproductive Biomedicine Online. 2012;24(5):485-6. doi: 10.1016/j.rbmo.2012.03.007, PMID 22578381.
8. Borgman CL, Furner J. Scholarly communication and bibliometrics. Annual Review of Information Science and Technology. 2002;36(1):2-72. doi: 10.1002/aris.1440360102.
9. Dolhey S. A bibliometric analysis of research on entrepreneurial intentions from 2000 to 2018. Journal of Research in Marketing and Entrepreneurship. 2019;21(2):180-99. doi: 10.1108/JRME-02-2019-0015.
10. Hu Y, Yu Z, Cheng X, Luo Y, Wen C. A bibliometric analysis and visualization of medical data mining research. Medicine. 2020;99(22):e20338. doi: 10.1097/MD.0000000000020338, PMID 32481411.
11. Duan L, Zhu G. Mapping theme trends and knowledge structure of magnetic resonance imaging studies of schizophrenia: A bibliometric analysis from 2004 to 2018. Frontiers in psychiatry. 2020;11:27. doi: 10.3389/fpsyt.2020.00027, PMID 32116844.
12. dos Santos BS, Steiner MT, Fenerich AT, Lima RH. Data mining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018. Computers & Industrial Engineering. 2019;138:106120. PMID 106120.
13. Tran BX, McIntyre RS, Latkin CA, Phan HT, Vu GT, Nguyen HLT, *et al.* The current research landscape on the artificial intelligence application in the management of depressive disorders: A bibliometric analysis. International journal of Environmental Research and Public Health. 2019;16(12):2150. doi: 10.3390/ijerph16122150.
14. Hew JJ. Hall of fame for mobile commerce and its applications: A bibliometric evaluation of a decade and a half (2000–2015). Telematics and Informatics. 2017;34(1):43-66. doi: 10.1016/j.tele.2016.04.003.
15. De Bakker FG, Groenewegen P, Den Hond F. A bibliometric analysis of 30 years of research and theory on corporate social responsibility and corporate social performance. Business & society. 2005;44(3):283-317. doi: 10.1177/0007650305278086.
16. Merigó JM, Mas-Tur A, Roig-Tierno N, Ribeiro-Soriano D. A bibliometric overview of the Journal of Business Research between 1973 and 2014. Journal of Business Research. 2015;68(12):2645-53. doi: 10.1016/j.jbusres.2015.04.006.
17. Li K, Rollins J, Yan E. Web of Science use in published research and review papers 1997–2017: A selective, dynamic, cross-domain, content-based analysis. Scientometrics. 2018;115(1):1-20. doi: 10.1007/s11192-017-2622-5, PMID 29527070.
18. Van Eck N, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 2010;84(2):523-38. doi: 10.1007/s11192-009-0146-3, PMID 20585380.
19. Aggarwal S, Singh R. Visual Exploration and Knowledge Discovery from Biomedical Dark Data. arXiv preprint arXiv:2009.13059. 2020.
20. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, Cannon TD, Krystal JH, Corlett PR. Cross-trial prediction of treatment

outcome in depression: a machine learning approach. The Lancet Psychiatry. 2016;3(3):243-50. doi: 10.1016/S2215-0366(15)00471-X, PMID 26803397.

21. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annual review of clinical psychology. 2017 May Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annual Review of Clinical Psychology. 2017;13:23-47. doi: 10.1146/annurev-clinpsy-032816-044949, PMID 28375728.
22. Hosseinifard B, Moradi MH, Rostami R. Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. Computer Methods and Programs in Biomedicine. 2013;109(3):339-45. doi: 10.1016/j.cmpb.2012.10.008, PMID 23122719.
23. Nouretdinov I, Costafreda SG, Gammerman A, Chervonenkis A, Vovk V, Vapnik V, Fu CH. Machine learning classification with confidence: application of transductive conformal predictors to MRI-based diagnostic and prognostic markers in depression. Neuroimage. 2011;56(2):809-13. doi: 10.1016/j.neuroimage.2010.05.023, PMID 20483379.
24. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, Rosen M, Ruef A, Dwyer DB, Paolini M, Chisholm K, Kambeitz J, Haidl T, Schmidt A. Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. JAMA Psychiatry. 2018;75(11):1156-72. doi: 10.1001/jamapsychiatry.2018.2165, PMID 30267047
25. Månsson KN, Frick A, Boraxbekk CJ, Marquand AF, Williams SC, Carlbring P, Andersson G, Furmark T. Predicting long-term outcome of Internet-delivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning. Translational Psychiatry. 2015;5(3):e530. doi: 10.1038/tp.2015.22, PMID 25781229.
26. Helbich M, Yao Y, Liu Y, Zhang J, Liu P, Wang R. Using deep learning to examine street view green and blue spaces and their associations with geriatric depression in Helbich M, Yao Y, Liu Y, Zhang J, Liu P, Wang R. Using deep learning to examine street view green and blue spaces and their associations with geriatric depression in Beijing, China. Environment International. 2019;126:107-17. doi: 10.1016/j.envint.2019.02.013, PMID 30797100.
27. Patel MJ, Khalaf A, Aizenstein HJ. Studying depression using imaging and machine learning methods. NeuroImage: Clinical. 2016;10:115-23. doi: 10.1016/j.nicl.2015.11.003, PMID 26759786.
28. Patel MJ, Andreescu C, Price JC, Edelman KL, Reynolds III CF, Aizenstein HJ. Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction. International Journal of Geriatric Psychiatry. 2015;30(10):1056-67. doi: 10.1002/gps.4262, PMID 25689482.
29. Lin E, Kuo PH, Liu YL, Yu YW, Yang AC, Tsai SJ. A deep learning approach for predicting antidepressant response in major depression using clinical and genetic biomarkers. Frontiers in Psychiatry. 2018;9:290. doi: 10.3389/fpsyt.2018.00290, PMID 30034349.
30. Galatzer-Levy IR, Ma S, Statnikov A, Yehuda R, Shalev AY. Utilization of machine learning for prediction of post-traumatic stress: a re-examination of cortisol in the prediction and pathways to non-remitting PTSD. Translational Psychiatry. 2017;7(3):e0. doi: 10.1038/tp.2017.38, PMID 28323285.
31. Oliveira L, Ladouceur CD, Phillips ML, Brammer M, Mourao-Miranda J. What does brain response to neutral faces tell us about major depression? Evidence from machine learning and fMRI. PLOS ONE. 2013;8(4):e60121. doi: 10.1371/journal.pone.0060121, PMID 23560073.
32. Lee Y, Ragguett RM, Mansur RB, Boutilier JJ, Rosenblat JD, Trevizol A, Brietzke E, Lin K, Pan Z, Subramaniapillai M, Chan TC. Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review. Journal of Affective Disorders. 2018;241:519-32. doi: 10.1016/j.jad.2018.08.073, PMID 30153635.
33. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJ, Dobson RJ, Dutta R. Characterisation of mental health conditions in social media using Informed Deep Learning. Scientific Reports. 2017;7(1):45141. doi: 10.1038/srep45141, PMID 28327593.
34. Michie S, Thomas J, Johnston M, Aonghusa PM, Shawe-Taylor J, Kelly MP, Deleris LA, Finnerty AN, Marques MM, Norris E, O'Mara-Eves A. The Human Behaviour-Change Project: Harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. Implementation Science. 2017;12(1):1-12.
35. Eck NJ, Waltman L. How to normalize cooccurrence data? An analysis of some well-known similarity measures. Journal of the American society for Information Science and Technology. 2009;60(8):1635-51. doi: 10.1002/asi.21075.
36. Laudano MC, Marzi G, Caputo A. A decade of the International Journal of Entrepreneurship and Small Business: A bibliometric analysis. International Journal of Entrepreneurship and Small Business. 2018;33(2):289-314. doi: 10.1504/IJESB.2018.090151.
37. Marshakova IV. System of document connections based on references. Nauchno-Tekhnicheskaya Informatsiya Seriya 2-Informatsionnye Protsessy I Sistemy. 1973;6:3-8.
38. Small H. Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for Information Science. 1973;24(4):265-9. doi: 10.1002/asi.4630240406.
39. Breiman L. Random forests. Machine learning. 2001;45(1):5-32. doi: 10.1023/A:1010933404324.
40. Kessler MM. Bibliographic coupling between scientific papers. American Documentation. 1963;14(1):10-25. doi: 10.1002/asi.5090140103.