ISSN No:-2456-2165

A Bibliometric Analysis of Heart Disease Detection using Artificial Intelligence Techniques: Trends, Influential Works, and Research Gaps

Tanushree Bharti
Department of CSE
Faculty of Computer Science and Engg.
Poornima University, Jaipur, Rajasthan

Pushpendra Kanwar
Department of CSE
Faculty of Computer Science and Engg.
Poornima University, Jaipur, Rajasthan

Abstract:- Advanced diagnostic techniques are required as cardiovascular diseases continue to pose a serious threat to global health. The scientific community has recently shown a great deal of interest in the application of deep learning techniques to the detection of heart disease. In order to synthesize the body of research on the use of deep learning in the detection of heart disease, this study provides a thorough bibliometric analysis. A wide variety of publications, including articles, conference papers, and reviews, are included in the analysis. These were obtained from Scopus and WoS databases. Total 662 documents are analyzed from these databases. The study looks at geographic distributions, historical trends, and influential figures in the field. We uncover key papers and authors through quantitative analyses, providing insight into the way research themes have changed over time. The study delves into co-authorship networks and institutional collaborations, offering valuable perspectives on the collaborative environment among scholars operating within this field. To find popular terms and hot topics, keyword analysis is used, which helps to provide a more sophisticated understanding of the main ideas guiding the research that is being done today.

Keywords:- *Bibliometric Analysis, Heart Diseases, Machine Learning, Deep Learning, Quantitative Analysis.*

I. INTRODUCTION

Cardiovascular diseases (CVDs) are the primary cause of morbidity and mortality worldwide, presenting a significant public health challenge. The necessity of developing precise and prompt diagnostic techniques has prompted a paradigm change in the field of heart disease detection towards the incorporation of deep learning approaches. The increasing amount of literature in this multidisciplinary field requires a methodical analysis in order to identify key trends, seminal works, and future research directions. This introduction sets the stage for a thorough bibliometric analysis by giving a thorough overview of the state of the art in deep learning-based heart disease detection.

This bibliometric analysis guides future research directions in addition to consolidating existing knowledge. Researchers have the opportunity to contribute to the ongoing development of deep learning applications in heart disease detection by examining the identified research gaps and emerging themes. As the field develops, cooperation between academics, organisations, and geographic areas will be essential for tackling the intricate problems related to cardiovascular health.

Citation trends demonstrated the recognition of foundational works and the impact and influence of important publications. For researchers who want to work with foundational concepts and methodologies, finding highly cited papers offers a starting point.

Essentially, this analysis adds to our collective knowledge of the state of the field and offers insightful information to researchers, clinicians, and policymakers alike. We can all work together to create novel approaches and enhanced diagnostic techniques by identifying the historical and contemporary research trajectories in deep learning-based heart disease detection. This will ultimately advance the larger objective of improving cardiovascular health globally.

A. *Significance of AI, ML, DL in Heart Disease Detection*

Artificial intelligence and machine learning play an important role in diagnosis of heart diseases nowadays but when it comes to the images these techniques have their limitations. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep neural networks (DNNs) are three examples of deep learning techniques that have proven to have exceptional powers in signal processing, image analysis, and feature extraction—all essential tasks in the context of cardiac diagnostics. These models offer a potential breakthrough in the early detection and prognostication of cardiovascular ailments by using hierarchical representations to automatically learn relevant features from raw data. To help researchers, clinicians, and policymakers fully utilise deep learning applications as the field develops, a comprehensive grasp of the scientific landscape is crucial.

B. *Research Objectives:*

Using machine learning and deep learning, this bibliometric analysis aims to untangle the complex web of research on heart disease detection. Our objectives are to map collaborative networks within the research community, identify recurrent themes, and evaluate the significance of important publications by methodically reviewing and quantitatively analysing a wide range of scholarly outputs. The knowledge gained from this analysis will be useful as a compass for determining future research directions in addition to providing a broad overview of the state of the field at the moment.

C. Structure of the Analysis:

The upcoming analysis covers a wide range of dimensions, such as geographical distributions, authorship networks, citation patterns, co-occurrence of keywords, and temporal trends. By using these lenses, we hope to offer a comprehensive understanding of the development and state of deep learning applications in the identification of heart disease, opening the door to well-informed choices in clinical practise and research. Section II describes the background and section III presents the methodology. Bibliometric analysis is presented in section IV.

II. BACKGROUND

The high incidence of cardiovascular diseases, which include heart failure, arrhythmias, and coronary artery disease, highlights the critical need for advanced diagnostic instruments. Even though they have their uses, traditional methods are not always accurate or efficient. A new era in medical diagnostics has been brought about in recent years by the development of deep learning, a subset of artificial intelligence (AI) distinguished by intricate neural network architectures. Deep learning models' ability to identify complex patterns in large datasets has great potential for improving the accuracy and speed of heart disease detection.

C A, Vidya [1] authored a review article concentrated on works published between 2017 and 2022. In addition, a comparative analysis is conducted to evaluate how well various deep-learning models identify cardiac disease. The advancements in deep learning applications for diagnosing heart disease and medical diagnosis are described in paper written by Bhandare and Trupti [2]. The evolution of the deep learning approach for medical data analysis is outlined, along with a review of the application, database, and learning system used in the automation process.

Ahsan and Md. Manjurul [3] presented a systematic literature review (SLR) approach to identify the problems related to unbalanced data in heart disease predictions in order to give a more comprehensive picture of the body of existing literature. Prior to that, they used 451 references that they had obtained from reputable journals between November 15, 2021, and 2012 to perform a meta-analysis. 49 pieces of cited literature have been carefully examined, with consideration given to the following aspects: kind of heart disease, algorithms, applications, and solutions.

The most frequently cited source was written by Alan S. Maisel [4] and appeared in the New England Journal of Medicine in 2002. Rahul C. Deo [5] wrote the second most-cited reference, which was published in Circulation in 2015. Author Jasvinder A. Singh's 2011 publication [6], included in the Cochrane Database of Systematic Reviews, is the third most cited reference. We looked at these top three articles. A. S. Maisel's most-cited reference from 2002 stated that B-type natriuretic peptide measurements significantly increased the independent predictive power of other clinical variables in models that used multiple logistic regression analysis to predict which patients had congestive heart failure. The second most frequently cited source by R. C. Deo from 2015 stated that one of its goals was to determine potential roadblocks to changing medical practise through statistical learning approaches and to talk about potential solutions. In the empirical Bayesian framework, the third most cited reference by J. A. Singh (2011) employed mixed-effects logistic regression with arm-based random-effects models. The aim was to aggregate the outcomes of biologics used in various conditions to derive critically important risk estimates.

III. METHODOLOGY

The information was obtained from Web of Science (WOS) and Scopus, two reputable and widely acknowledged repositories. Due to their clarity in indexing peer-reviewed journals and quality, both databases are well-liked. Query 1 shows the search and selection criteria for the bibliometric analysis. An initial search was conducted using the title, abstract, and keyword of the publication, yielding 13,375 articles in the Scopus database. The keywords machine learning, disease diagnosis, disease, and disease prediction were taken into consideration.

The number of articles was reduced to 874 after the search was restricted to the English-language, peer-reviewed, open-access literature published between 2017 and 2023. After that, we narrow down the term to terms like "subject area", "article type", "document type", etc. Lastly, we performed bibliometric analyses on 662 articles that we had chosen. The following query was given in the input to extract the articles.

Query-1 = ((Heart OR cardiovascular OR CVD) AND (Disease OR Diseases) AND (Detection OR Diagnosis) AND (Machine Learning OR Deep Learning OR AI OR Artificial Intelligence))

For refining the results the following query was given in the advanced search option of Scopus database.

TITLE (("Heart disease" OR "Heart diseases" OR "cardiovascular disease" OR "cardiovascular diseases" OR cvd) AND (detection OR diagnosis OR prediction) AND ("Machine Learning" OR "Deep Learning" OR ai OR ml OR dl OR "Artificial Intelligence")) AND PUBYEAR > 2016 AND PUBYEAR < 2025 AND (LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "ENGI") OR LIMIT-TO (SUBJAREA , "MATH") OR LIMIT-TO (SUBJAREA , "DECI")) AND (LIMIT-TO (DOCTYPE , "re") OR LIMIT-TO (DOCTYPE , "ch") OR LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp")) AND (LIMIT-TO (LANGUAGE , "English"))

ISSN No:-2456-2165

Biblioshiny of RStudio and VOSViewer tools were used to perform the bibliometric analysis of Scopus and WoS databases. The file type is Bibtex for Scopus database and CSV for WoS database.

IV. BIBLIOMETRIC ANALYSIS

A. Main information about data

Table 1 shows a main information about the metadata where total 662 articles were extracted from the database. The time span is from 2017 to 2023. Total author's keywords are 1007 and total authors are 1588. The annual growth rate shows the notable increase in research article publications over the previous ten years, demonstrating the increased interest in machine learning-based disease diagnosis among academic communities.

It is evident from the table that there was comparatively little published articles by single authors. The per document authorship rate is 3.99. There were only 31 review articles and rest are research articles. We anticipate that the publication of ML-based disease diagnosis will grow exponentially in 2022 and the years that follow.

TABLE 1. MAIN INFORMATION ABOUT DATA

Description	**Results**
Timespan	2017:2023
Sources (Journals, Books, etc)	264
Documents	662
Annual Growth Rate %	44.22
Document Average Age	2.03
Average citations per doc	6.646
References	15301
Keywords Plus (ID)	2044
Author's Keywords (DE)	1007
Authors	1588
Single-authored docs	10
Co-Authors per Doc	3.99
International co-authorships %	15.99
Article	192
Conference paper	221
Review	31

B. Publications by citations

As it gathers data and concepts about the major writers of this domain, the study assesses the total number of citations. The top 10 articles that have been cited the most, as compiled from the Scopus and WOS databases, are shown in Fig. 1 below. Every author in Fig. 1 received a citation in the range of 100 to 2500. Be aware that because different databases may use different indexing strategies and time periods, the total number of citations may differ from citations found in Google Scholar and other databases. Fig. 1 shows that author Mohan's work received the highest number of citations.

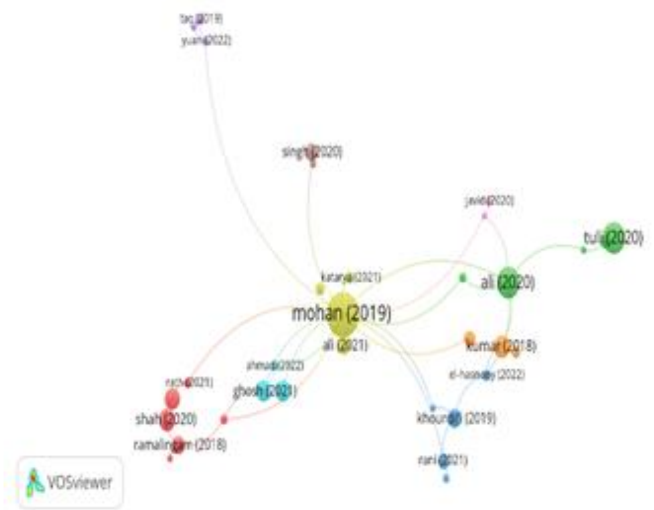

Fig. 1 The Most cited articles [Source VOSviewer]

C. Most Productive Countries by Citation

Fig. 2 shows the ten most productive nations. After examining the top nations, we discovered that the United States, China, and India are the most significant nations in terms of scientific output. Fig. 2 shows that the India and China approximate same number of citations. United Kingdom and Canada are followed by these above countries.

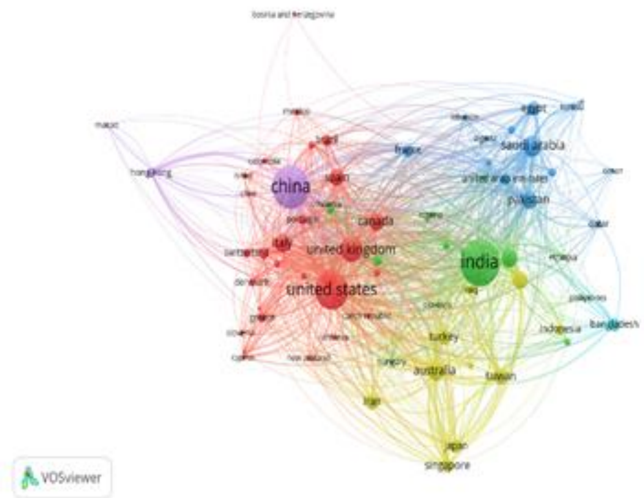

Fig. 2. The most productive countries. [Source VOSviewer]

D. Publication by countries

The top 20 nations or regions for heart disease detection in this domain are shown in Fig. 3. This figure shows published work in top 20 countries with size of circle, larger size shows more number of articles from 2017 to 2023. The comprehensive documents are produced in collaboration with other nations as well as those from a single nation. Based on the number of publications, India perform better than any other country, indicating that they started their research endeavours earlier than the majority of other nations in the world.

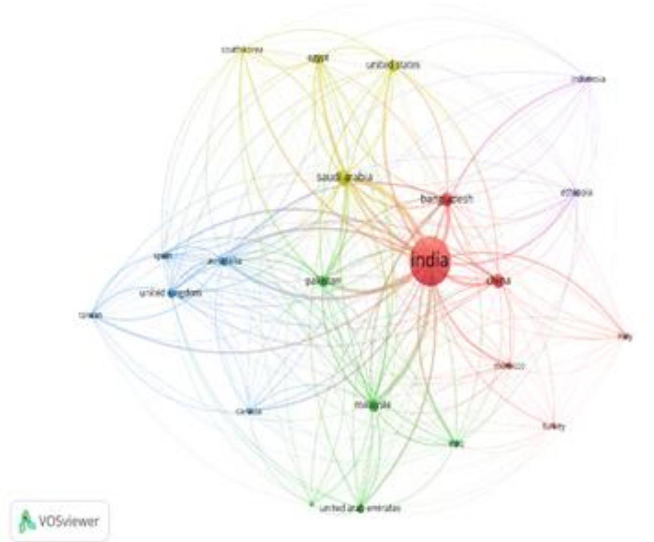

Fig. 3 Publications by Countries. [Source VOSviewer]

E. Publication by institutions

The department of electronics and biomedical is the most productive academic institution in terms of published papers, as indicated in Fig. 4, based on our analysis of the top ten universities. Universities in the United States, including Harvard Medical School and University of California, San Francisco, have been revealed to be among the top three most productive, having published 46 and 43 papers respectively. The Figure shows that the USA accounts for seven of the ten most productive institutions, with an active focus on heart disease detection research.

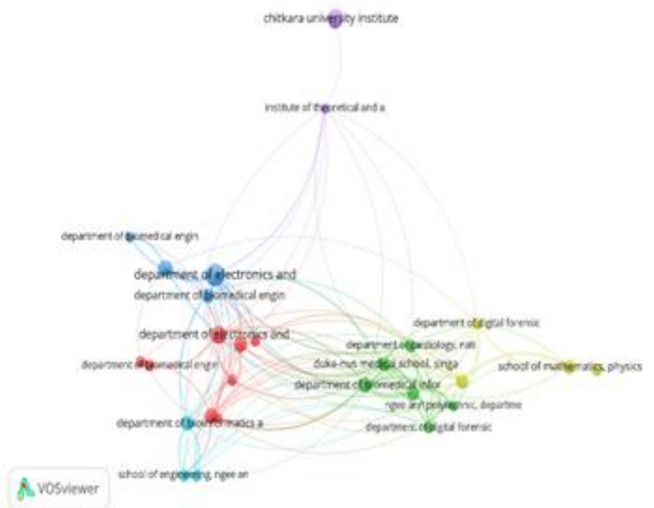

Fig. 4. Publication by Institutions [Source VOSviewer]

F. Common words in title and abstract

VOSviewer software was used to determine the most commonly used terms in titles and abstracts, as shown in Fig. 5. "Machine learning" is the most frequently used term in title-abstracts, "Artificial Intelligence" is next appearing "feature selection" and so on. Even though heart diseases detection is the focus of most of the articles, some of the most important terms—like "disease" and "algorithms"—are uncommon and make up less than 2% of all the words that the authors use most frequently. Fig. 6 shows the cluster of authors who did work in this domain.

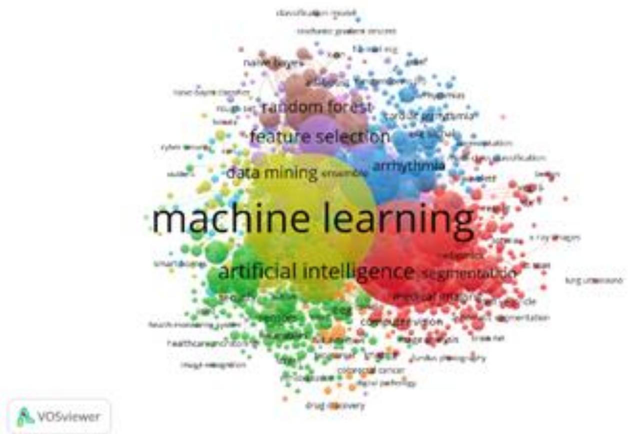

Fig. 5 Common Keywords in Articles [Source VOSviewer]

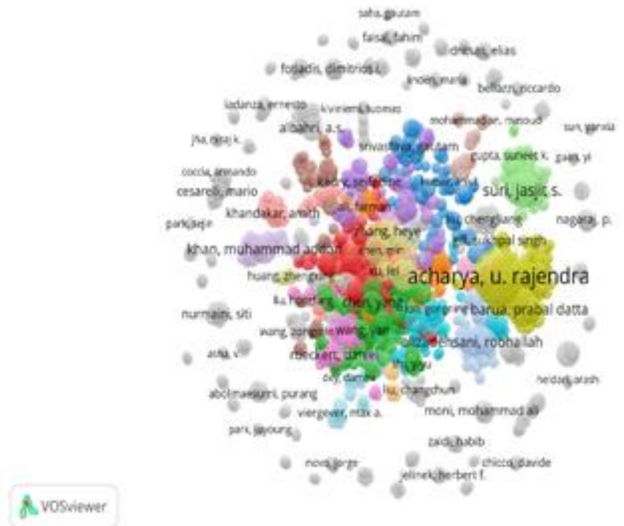

Fig. 6. Cluster of Authors [Source VOSviewer]

V. RESULT AND DISCUSSION

A. Observations and findings

The insights and outcomes of the metadata analysis are presented in the section IV. Based on a metadata analysis of 662 chosen papers and a content analysis of 15 publications, the conclusions were drawn. More papers are being published from India but these are not getting cited so less productivity. Scopus database have more publication as compared to the WoS database. More number of articles are published from India as compared with other countries.

An extensive analysis is conducted using 662 articles of literature to gain an understanding of the methods and trends currently employed in the diagnosis of heart disease. The type of heart disease, applications, machine learning algorithms, and imbalance solution were all taken into account in the comprehensive analysis. Mostly work is done in machine learning domain.

B. Summary of Analysis

By analyzing the words and keywords in the title, abstract, and keywords it can be conclude that less work was done as compared to the machine learning in the field of deep learning. The frequencies of the words CNN, RNN, Autoecoders, CT images, MRI images, and ECG data are very less as compared to the other data mining related words. More work can be done on image de-noising using Auto-encoders for improving the performance of CNN.

VI. CONCLUSION

The bibliometric analysis of deep learning-based heart disease detection offers a thorough picture of the changing field at this crucial nexus of artificial intelligence and cardiovascular health. A wide range of scholarly outputs have been synthesized, revealing important trends, seminal works, and cooperative networks in the research community. The temporal analysis showed a consistent increase in publications, highlighting the growing awareness of machine learning and deep learning's potential to transform cardiac diagnostics. Geographic distributions revealed a global participation, with contributions coming from various regions, indicating that this field is widely acknowledged for its importance.

The intricate webs of cooperation among researchers and institutions were revealed by authorship networks and collaboration patterns. The discovery of prolific writers and cooperative groups not only recognizes the collaborative aspect of scientific development but also points to potential areas for future interdisciplinary collaboration to advance the field. The analysis of keywords offered valuable perspectives on the terms that are changing in the field, demonstrating the fluidity of research priorities. Researchers' shifting interests are reflected in the emergence of particular terms, and the discovery of keywords with enduring relevance provides a road map for understanding the fundamental ideas of deep learning-based heart disease detection.

REFERENCES

[1]. C A, Vidya and V. Baby Shalini. “Systematic Review on Deep Learning-based Heart Disease Diagnosis.” *2023 2nd International Conference on Edge Computing and Applications (ICECAA)* (2023): 908-912.

[2]. Bhandare, Trupti Vasantrao and Selvarani Rangasamy. “Review on Heart Disease Diagnosis Using Deep Learning Methods.” *Int. J. Next Gener. Comput.* 12 (2021): n. pag.

[3]. Ahsan, Md. Manjurul and Zahed Siddique. “Machine Learning-Based Heart Disease Diagnosis: A Systematic Literature Review.” *Artificial intelligence in medicine* 128 (2021): 102289 .

[4]. Maisel AS, Krishnaswamy P, Nowak RM, McCord J, Hollander JE, Duc P,et al. Rapid measurement of B-type natriuretic peptide in the emergencydiagnosis of heart failure. New Engl J Med. (2002)347:161 doi:10.1056/NEJMoa020233

[5]. Deo RC. Machine learning in medicine. Circulation. (2015) 132:1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593

[6]. Singh JA, Wells GA, Christensen R, Ghogomu TE, Maxwell L, MacDonald JK,et al.. Adverse effects of biologics: a network meta-analysis and Cochrane overview.Cochrane Database Syst Rev. (2011) 2:14651858. doi: 10.1002/14651858.CD008794.pub2

[7]. Md Manjurul Ahsan, Zahed Siddique, Machine learning-based heart disease diagnosis: A systematic literature review, Artificial Intelligence in Medicine, Volume 128, 2022, 102289, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2022.102289.

[8]. Y. Singh, M. Saini and Savita, "Impact and Performance Analysis of Various Activation Functions for Classification Problems," 2023 IEEE International Conference on Contemporary Computing and Communications (InC4), Bangalore, India, 2023, pp. 1-7, doi: 10.1109/InC457730.2023.10263129.

[9]. Vispute, S., Saini, M.L. (2022). Automation in Agriculture: A Systematic Survey of Research Activities in Agriculture Decision Support Systems Using Machine Learning. In: Singh, P.K., Wierzchoń, S.T., Chhabra, J.K., Tanwar, S. (eds) Futuristic Trends in Networks and Computing Technologies . Lecture Notes in Electrical Engineering, vol 936. Springer, Singapore. https://doi.org/10.1007/978-981-19-5037-7_56

[10]. M. L. Saini, V. K. Sharma and A. Kumar, "An Efficient Single and Double Error Correcting Block Codes with Low Redundancy for Digital Communications," 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 2023, pp. 827-831, doi: 10.1109/InCACCT57535.2023.10141727.

[11]. Vispute, S.; Saini, M.L. Performance Analysis of Soil Health Classifiers Using Data Analytics Tools and Techniques for Best Model and Tool Selection. Int. J. Online Biomed. Eng. 2022, 18, 169–189.

[12]. Shailaja Pede and Madan Lal Saini, "A Brief Bibliometric Analysis and Visualisation of Scopus and WoS databases on Blockchain Technology in Healthcare Domain", Library Philosophy and Practice (e-journal), vol. 5106, 2021.

[13]. M. Gehlot and M. L. Saini, "Analysis of Different CNN Architectures for Tomato Leaf Disease Classification," 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, 2020, pp. 1-6, doi: 10.1109/ICRAIE51050.2020.9358279.

[14]. Varun Sapra, Madan Lal Saini, “Deep learning network for identification of Ischemia using clinical data”, International Journal of Engineering and Advanced Technology, ISSN: 2249-8958, Volume-8 Issue-5, June 2019.

[15]. Varun Sapra, Madan Lal Saini, “Computational Intelligence for Detection of Coronary Artery Disease with Optimized Features”, International Journal of Innovative Technology and Exploring Engineering, ISSN: 2278-3075 Volume 8, Issue-6C, Pages 144-148, April 2019.

[16]. S. Kulshrestha and M. L. Saini, "Study for the Prediction of E-Commerce Business Market Growth using Machine Learning Algorithm," *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, Jaipur, India, 2020, pp. 1-6, doi: 10.1109/ICRAIE51050.2020.9358275

[17]. Sapra Varun , Saini M.L and Verma Luxmi, Identification of Coronary Artery Disease using Artificial Neural Network and Case-Based Reasoning, Recent Advances in Computer Science and Communications 2021; 14(8) . https://dx.doi.org/10.2174/266625581399920061322540 4

[18]. M. L. Saini, J. Panduro-Ramirez, J. Padilla-Caballero, A. Saxena, M. Tiwari and K. Ravi, "A Study on the Potential Role of Blockchain in Future Wireless Mobile Networks," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 547-550, doi: 10.1109/ICACITE57410.2023.10182375.

[19]. Kavita Lal, Madan Lal Saini; A study on deep fake identification techniques using deep learning. AIP Conf. Proc. 15 June 2023; 2782 (1): 020155. https://doi.org/10.1063/5.0154828

[20]. dos Santos, Bruno Samways, et al. "Data mining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018." *Computers & Industrial Engineering* 138 (2019): 106120.

[21]. Kuang, Xuyuan, et al. "Bibliometric analysis of 100 top cited articles of heart failure–associated diseases in combination with machine learning." *Frontiers in Cardiovascular Medicine* 10 (2023): 1158509.

[22]. Guo, Yuqi, et al. "Artificial intelligence in health care: bibliometric analysis." *Journal of Medical Internet Research* 22.7 (2020): e18228.