



A Statistical Presentation of Heart Disease Prediction using Machine and Deep Learning Perspectives

Mohammad Amanulla^{1,*}, G Sanjiv Rao^{2,b}, A Obulesu^{3,c} and B Srinivas^{1,d}

¹Assistant Professor, Aditya College of Engineering & Technology, Surampalem, Andhra Pradesh, India.

²Professor, Aditya College of Engineering & Technology, Surampalem, Andhra Pradesh, India.

³Associate Professor, Vidya Jyothi Institute of Technology, Hyderabad, Telangana, India.

^ae-mail: Avuku06@gmail.com, ^be-mail: sanjiv.rao@acet.ac.in, ^ce-mail: b.srinivas44@gmail.com

^{*}Corresponding author's e-mail: amanmdphd@gmail.com

To Cite this Article

Mohammad Amanulla, G Sanjiv Rao, A Obulesu and B Srinivas. Supervised Dynamic Pattern Evaluation using Hybrid Fuzzy C-means Clustering for Handling Multi Dimensional Data. International Journal for Modern Trends in Science and Technology 2023, 9(07), pp. 106-115. <https://doi.org/10.46501/IJMTST0907018>

Article Info

Received: 02 April 2023; Accepted: 13 April 2023; Published: 19 July 2023.

Copyright © 2023 Mohammad Amanulla et al. This is an open access article distributed under the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: This analysis aims at heart disease prediction and diagnosis using machine learning, deep learning techniques from 2012 to 2021, using statistical methods.

Methods: Retrieved various heart disease prediction articles from popular databases like Scopus, from 2012 to 2021 research articles are considered to analyse. To do and receive results like documents by affiliation, type, sponsors and so on, scopus analyser is used. As far as network analysis is concerned VOSviewer Version 1.6.17 is used to show the analysis relations among citation, occurrences and co-authorship etc.

Results: On heart disease prediction, the database results 717 articles to study from 2012 to 2021. India has contributed maximum articles from 2012 to 2021 is shown from statistical and network analysis. Different parameters of network analysis are the evidence to show the subject's potential in the field of research.

Conclusions: A huge scope is expected to contribute in future research in areas like NN (neural network advanced algorithms, DL (deep learning), and ML (machine learning) is shown from different parameters of network analysis. English has the best number, a total of 717 articles are resulted from the search of scopus keyword. The potential in the topic is shown from statistical analysis of authors, documents, affiliations and country.

KEYWORDS: heart disease, disease prediction, machine learning, deep learning.

1. INTRODUCTION

Heart disease is only the main cause of morbidity and mortality globally: it accounts for more deaths annually than the other cause. In 2019 according to the sources of

WHO, a total of 17.9 million people died because of this Coronary Vascular Disease. It is shown as 32% of globally occurred deaths. Heart failure and heart stroke [1] deaths covers 85%. Over 3/4 of those deaths occurred

in low-and socio-economic class nations. Coronary cardiovascular disease (also known as heart attack) is far and away the foremost fatal and therefore the commonest, over all heart diseases, within the US, for instance, it's calculated that every 40 seconds someone experiencing an attack and about 805,000 Americans have an attack per annum according to CDC 2019.

The term "coronary illness" refers to many sorts of unhealthy heart conditions. the foremost common sort of heart condition within the u. s. is the flow of blood to the heart is affected by coronary artery disease which is also known as CAD, the heart attack is caused due to decreased blood flow to the heart. Sometimes heart condition could also be not assessed and not diagnosed until an individual experiences symptoms of an arrhythmia, heart attack or coronary failure. Heart attack symptoms include: severe heart pain, dizziness, indigestion, extreme fatigue, neck pain, vomiting (nausea), heartburn. Reduces the breath in neck veins, abdomen, feet swelling, ankles, legs and also makes fatigue, when patient suffers with heart attack. Smoking, high blood cholesterol, Diabetes, Excessive alcohol use, Overweight and obesity are the primary causes of cardio vascular diseases (CVD).

Physical exam, family medical record, regular habits are best to diagnose and detect one's heart condition. For heart condition diagnosis we rely on ECG(Electro Cardio Gram)[2][3]used in blood tests, Stress Cardio Gram[16], Holter monitoring, Echocardiogram, Cardiac catheterization[4][5][6], CT Scan(Computerized Tomography)[7]. To collect images of heart Magnetic Resonance Imaging (MRI): A detailed image of our heart can be generated from magnetic field and ratio waves generated from computer by a cardiac MRI. This detection and diagnosis is incredibly critical and it requires a trained and experienced cardiologist to conclude about the diagnosis. Hence it's always necessary to produce an automatic system for this diagnosis. Doctors and scientists hopefully turned to machine learning (ML) techniques [9][10][11]to develop tools of screening and this can be due to their experience in pattern recognition and classification when contrast with other conventional statistical approaches.

Different features are extracted from the physical examination and health diagnostic reports for the reason of selecting classification models. These include age, sex, habits, chronic diseases, earlier health reports, BP

reports, Obesity, alcohol consumption frequency etc.After selecting the specified features the classifiers are accustomed to detect the probability of heart condition. Different classifiers that are used for the detection purpose are Logistic regression, K-Nearest Neighbours [12], Support Vector Machine [14][15], decision tree, random forest, convolutional neural network (CNN) [13]and Naïve Bayes. Additionally to those algorithms other researchers also initiated the utilization of varied bio-inspired optimization algorithms.

2. FUNDAMENTAL METHODS

2.1 Primary Search on Scopus Database

Scopus, ScienceDirect, Web of Science, IEEE Xplore etc., are worldwide popular databases for research. Really these databases having publications in extensive range, Scopus is the prominent databases among all above databases listed, analysis is initiated with the help of scopus database. Total 717 publications resulted from the keywords used in this search. There is no constraint in terms of language, country, type of document etc applied worldwide to search database using various keywords. Information like source, author, country, documents, citations are associated with each publications is used for analysis. Basic keywords used in search

Table 1. Primary and Secondary Keywords used in search.

Fundamental Keyword	Search	Heart Prediction habits.	Disease by
Primary keywords	Search includes (AND)	Heart disease AND Prediction AND using AND machine learning.	AND
Secondary Keywords	Search includes (OR)	Machine learning OR Deep learning	

Query used to search Scopus database documents as follows:

TITLE-ABS-KEY (heart AND disease AND prediction AND using AND machine AND learning OR deep AND learning) AND (LIMIT-TO (SUBJAREA ,

"COMP")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012))

2.2 Initial Search Outcomes

Publications are retrieved from scopus database using different keywords related to heart disease prediction. The result of this analysis is depends on the language. Out of 717 publications, 716 are in English and 1 in Turkish.

Table 2. Publications count according to language

Language	Number of publications
English	716
Turkish	1

Source: <https://www.scopus.com/results> (dated 02nd Sept. 2021)

2.3 Top 15 keywords used in publications

Machine learning keyword secured first place in publications with the highest in number. List of top 15 keywords are shown in the table 3. These keywords identified in addition to basic keywords listed above. All the listed keywords in the table are associated with health and technology.

Table 3: keywords used in search results with the highest publications

S No	Keywords	Publications
1	Machine Learning	339
2	Diseases	308
3	Cardiology	271
4	Heart Disease	231
5	Heart	206
6	Forecasting	196
7	Learning Systems	168
8	Diagnosis	134
9	Decision Trees	114
10	Learning Algorithms	106
11	Support Vector Machines	90
12	Predictive Analysis	82
13	Classification(of Information)	80
14	Blood Pressure	28

Source: <https://www.scopus.com/results> (dated 02nd Sept. 2021)

3. PERFORMANCE ANALYSIS

In addition to scopus database analyse, VOSviewer 1.6.17 software version is used for network analysis. It has provided a new perception to do research in terms of co-occurrence, co-citation and bibliometric coupling etc.

Analysis performed on Documents as follows.

Statistical Analysis of Database by Documents

1. Affiliation analysis
2. Subject area analysis
3. Source analysis
4. Author analysis
5. Year analysis
6. Type analysis
7. Country analysis
8. Funding agencies analysis

Network Analysis of Database by Documents

1. CA (Co Authorization): Authors, organizations, country
2. CO (Co- Occurrence): All keywords, Author keywords, Index keywords
3. CAs (Citation Analysis): Sources, authors, organizations, country
4. BC (Bibliographic coupling): Documents, Authors

4. RESULTS AND DISCUSSION

4.1 Statistical Analysis

4.1.1 Analysis by Affiliations

Analysis is carried out on top 10 affiliations. It is found that, Vellore Institute of Technology has more than half of the affiliations followed by K L Deemed to be University, all affiliations secured more than 5 publications in the field of heart disease is shown in figure 1.

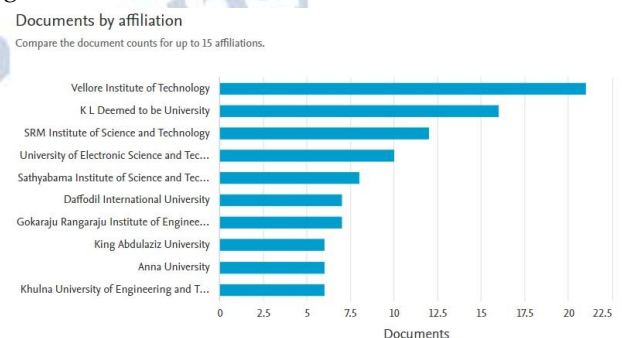


Figure 1. Analysis by affiliation

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

4.1.2 Analysis by Subject Area

Through Physical examination and blood trails of different cholesterol in our blood can possibly predict heart diseases. These are tried utilizing testing kits available in the market such are approved by FDA. Highest publications are contributed in this subject area are from computer science engineering (41.1%), only engineering (19.2%) and remaining fields comes under 39.7% is shown in figure 2.

Documents by subject area

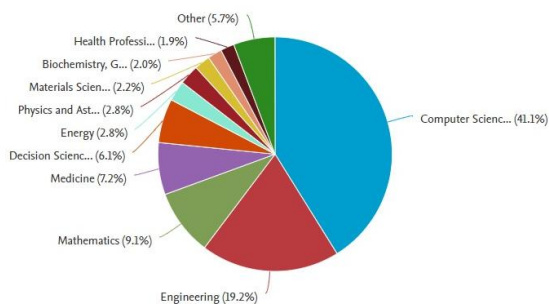


Figure 2. Analysis by Subject area

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

4.1.3 Analysis by Sources

The following figure 3 shows year-wise publication statistics of sources. Graphical representation is generated by network analysis software that shows the number of published documents in chronological order of considered scopus database. Conferences, book chapters, reviews, journals, etc., are considered as sources.

Documents per year by source

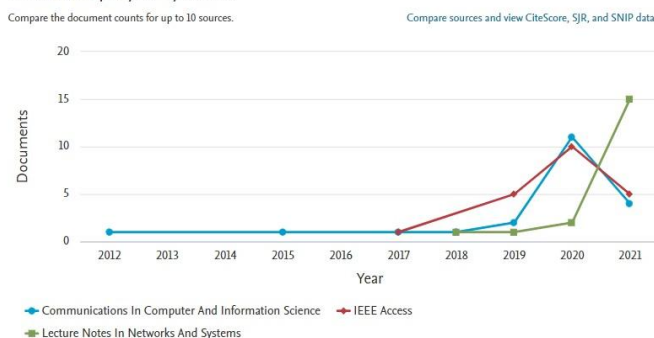


Figure 3. Analysis by Sources

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

4.1.4 Analysis by Author

Statistical Analysis is performed by author on scopus database and displays a list of top 10 is shown in figure 4. Most of the authors have published documents

from 4 to 5 ranges. From the table it is clearly visible that Raihan. M has maximum publications i.e. 6.

Documents by author

Compare the document counts for up to 15 authors.

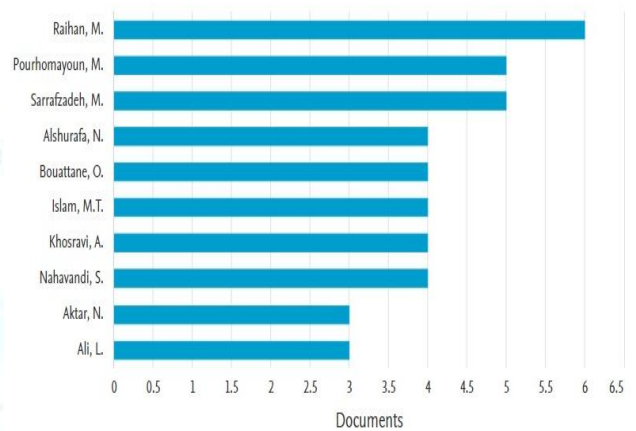


Figure 4. Analysis by author

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

4.1.5 Analysis by Year

The following table 4 shows year-wise publication statistics of sources in each individual year. Graphical representation is generated by network analysis software that shows the number of published documents, in chronological order of considered scopus database, is retrieved from 2012 to 2021. These representations are the evidence to shows the maximum publications possible during 2020 and 2021. It is proved that huge scope is possible in this area. Conferences, book chapters, reviews, journals, etc., are considered as sources is shown in figure 5.

Table 4: Publication statistics by Year

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

Year	Publications
2012	10
2013	12
2014	11
2015	20
2016	22
2017	31
2018	62
2019	168
2020	209
2021	172

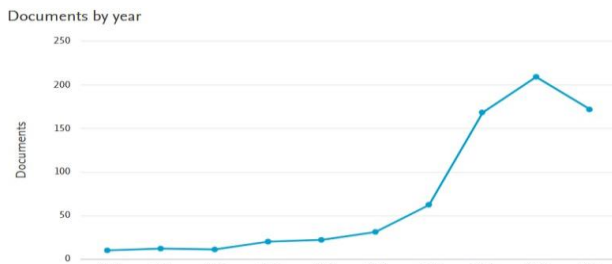


Figure 5. Analysis by Year

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

4.1.6 Analysis by Type

The table 5 below clearly shows that conferences, Articles have the maximum number of publications than other document types that are considered from the database is shown in figure 6.

Table 5: Analysis by Document Types

S No	Type of Document	Publications
1.	Conference Paper	346
2.	Article	303
3.	Conference Review	35
4.	Book Chapter	21
5.	Review	11
6.	Note	1
Total		717

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

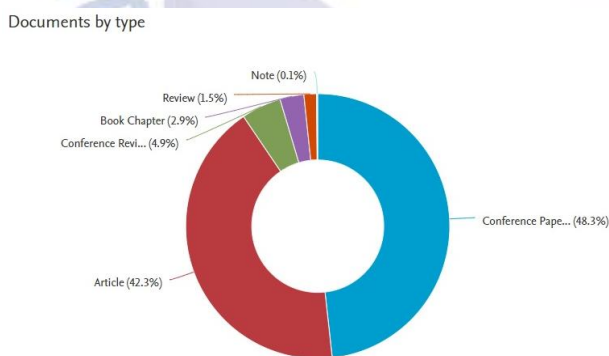


Figure 6. Analysis by Type

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

4.1.7 Analysis by Country or Territory

India is having maximum number of documents published in terms of countries when analysis is performed on scopus database. Statistical analysis also shows United States and China are in second and third in the below figure 7.

Documents by country or territory
Compare the document counts for up to 15 countries/territories.

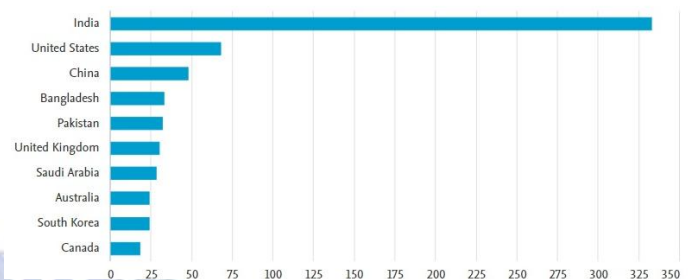


Figure 7. Analysis by country or territory

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

4.1.8 Analysis by Funding Sponsors

Health science Institutions contributed more in funding sponsors, is clearly shown in the below figure 8. China has occupied first place by funding maximum to the National Nature Science.

Documents by funding sponsor
Compare the document counts for up to 15 funding sponsors.

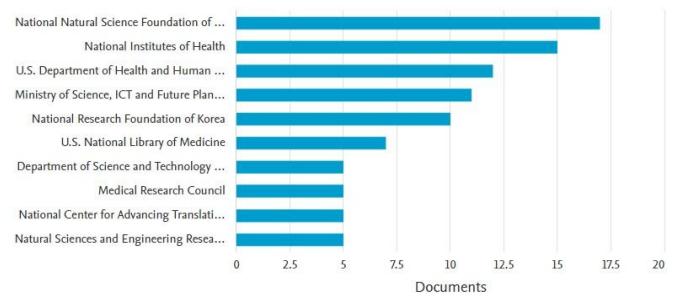


Figure 8. Analysis by funding sponsor

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

4.2 Network Analysis

4.2.1 Co-authorship

A) Co-authorship and Authors

Performed analysis by limiting to 03 distinct parameters like authors, countries and organizations are considered. Set a threshold of 3 as minimum number of documents and number of authors are set to 25.

The result figure 9 shows that 42 authors met the threshold out of 2068 authors. Co-authorship link strength is obtained highest for Raihan M with highest link strength of 15 of a total citations 09. The figure shows a total of 14 authors have co-authorship relation. Analysis input is shown below.

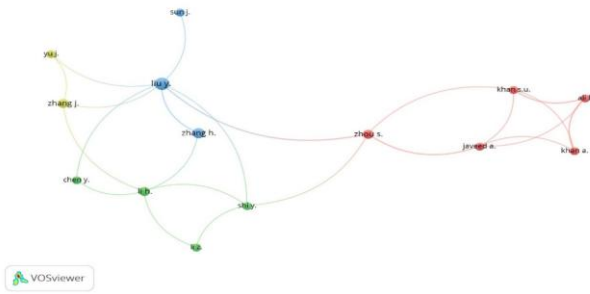


Figure 9. Co-authorship Network Analysis in Terms of Authors

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

B) Co-authorship and Organizations

Analysis is carried out by considering a limit of 02 documents per each organization, to show the relation between co-authorship and organizations. From the analysis it is observed that 28 organizations met the threshold out of 1340 documents. It is found that 4 organizations are with 6 as highest link strength. In this regard College of agriculture, samawah, Iraq has highest citations as 14 is shown in figure 10.

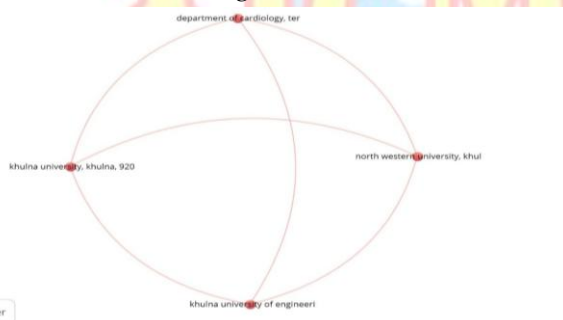


Figure 10. Co-authorship and Organizations Network analysis

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

C) Co-authorship and Country

Co-authorship and country relation is obtained as below. As a result, out of 77 countries 29 countries met a minimum threshold of 5 in each country. With 264 as highest citations and 47 as link strength Pakistan is in first in the list is shown in figure 11. In the highest documents category, with 304 India got the highest.

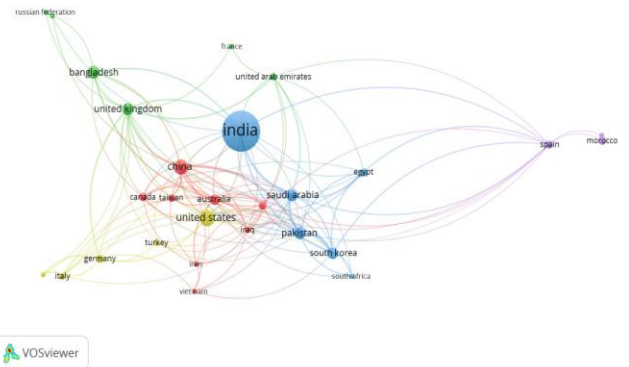


Figure 11. Co-authorship and Country Network analysis

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

4.2.2 Network Analysis of Co-occurrences;

A) Co-occurrence and all keywords

According to the following analysis “Machine Learning”, “Disease” and “Cardiology” keywords used maximum in all the obtained documents. As a limit of 5 occurrences, 368 out of 4160 are met the given constraint is shown in figure 12.

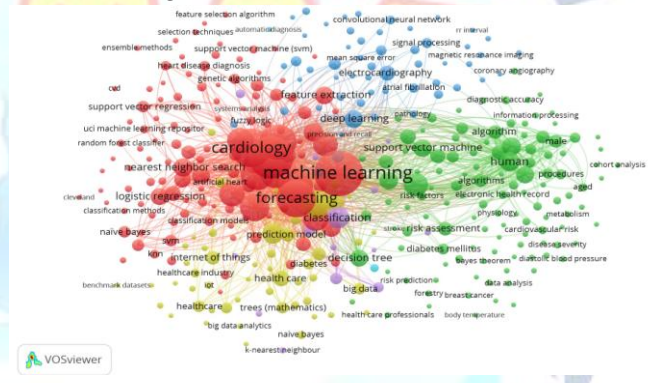


Figure 12. Co-occurrence and keyword Network analysis

<https://www.scopus.com/results> (dated 02nd Sept. 2021)

B) Co-occurrence and Author keywords

From the analysis of Co-occurrence and Author keyword relation, 88 keywords out of 1422 met the threshold by limiting it to 5 per author is shown from figure 13. Machine learning have highest with 283 with a link strength of 634 followed by heart disease and classification.

publications. The result of the search is only the evidence that India has the highest publications among all countries. In the subject area category Computer Science Engineering occupied a major portion of 41.1% among all other categories. Among 717 documents 716 are in English language and 1 is in Turkish. In the category of document type, conference paper, articles are in order to occupy the maximum number. These results are obtained by considering different arguments in search on the database. Network analysis is carried out on scopus database. The network analysis justifies the significance of co-authorship, citation, bc(bibliographic coupling) and, co-occurrence analysis on the considered scopus database. VOSViewer 1.6.17 software has been used for network analysis. It also shows that effective concentration on heart disease prediction is going on in these 2020 and 2021 years. This analysis is expecting a huge research scope in upcoming years.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

- [1] P. R. Savarapu, M. Shankar, S. Itnal, R. K. Madupu, C. Z. Basha and E. Sreedevi, "Advanced prediction of Heart diseases using Artificial Neural Network and Genetic Algorithm," *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 875-878, doi: 10.1109/ICCMC51019.2021.9418459.
- [2] T. G. Keshavamurthy and M. N. Eshwarappa, "Review paper on denoising of ECG signal," *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 2017, pp. 1-4, doi: 10.1109/ICECCT.2017.8117941.
- [3] B. Yavorskyy, M. Bachynskyy and V. Falendysh, "Provide of confidence for mobile Holter monitoring of heart rhythm variability," *2009 10th International Conference - The Experience of Designing and Application of CAD Systems in Microelectronics*, 2009, pp. 171-173.
- [4] B. Suelze *et al.*, "Waving at the Heart: Implementation of a Kinect-based real-time interactive control system for viewing cineangiogram loops during cardiac catheterization procedures," *Computing in Cardiology 2013*, 2013, pp. 229-232.
- [5] L. Sun *et al.*, "Direct measurement of a patient's entrance skin dose during pediatric cardiac catheterization," in *Journal of Radiation Research*, vol. 55, no. 6, pp. 1122-1130, Nov. 2014, doi: 10.1093/jrr/rru050.
- [6] O. A. Alim and M. Shoura, "A new method for dynamic road mapping of coronary angiography cardiac catheterization cine loops," *Proceedings of the Twentieth National Radio Science Conference (NRSC'2003) (IEEE Cat. No.03EX665)*, 2003, pp. K4-1, doi: 10.1109/NRSC.2003.157360.
- [7] F. Tvard, A. Simon, C. Leclercq, P. Mabo, A. Hernandez and M. Garreau, "Data fusion of left ventricle Electro-Anatomic Mapping and Multislice Computerized Tomography for Cardiac Resynchronisation Therapy optimization," *2009 36th Annual Computers in Cardiology Conference (CinC)*, 2009, pp. 613-616.
- [8] Yudong Zhu, M. Drangova and N. J. Pelc, "Estimation of deformation gradient and strain from cine-PC velocity data [cardiac magnetic resonance imaging]," in *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 840-851, Dec. 1997, doi: 10.1109/42.650880.
- [9] C. -H. Lin, P. -K. Yang, Y. -C. Lin and P. -K. Fu, "On Machine Learning Models for Heart Disease Diagnosis," *2020 IEEE 2nd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, 2020, pp. 158-161, doi: 10.1109/ECBIOS50299.2020.9203614.
- [10] S. Ismaeel, A. Miri and D. Chourishi, "Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis," *2015 IEEE Canada International Humanitarian Technology Conference (IHTC2015)*, 2015, pp. 1-3, doi: 10.1109/IHTC.2015.7238043.
- [11] A. Erdoğan and S. Güney, "Heart Disease Prediction by Using Machine Learning Algorithms," *2020 28th Signal Processing and Communications Applications Conference (SIU)*, 2020, pp. 1-4, doi: 10.1109/SIU49456.2020.9302468.
- [12] P. Sujatha and K. Mahalakshmi, "Performance Evaluation of Supervised Machine Learning Algorithms in Prediction of Heart Disease," *2020 IEEE International Conference for Innovation in Technology (INOCON)*, 2020, pp. 1-7, doi: 10.1109/INOCON50539.2020.9298354.
- [13] W. Zhang and J. Han, "Towards heart sound classification without segmentation using convolutional neural network," *2017 Computing in Cardiology (CinC)*, 2017, pp. 1-4, doi: 10.22489/CinC.2017.254-164.
- [14] T. -N. Nguyen, T. -H. Nguyen, D. -D. Vo and T. -D. Nguyen, "Multi-class Support Vector Machine Algorithm for Heart Disease Classification," *2020 5th International Conference on Green Technology and Sustainable Development (GTSD)*, 2020, pp. 137-140, doi: 10.1109/GTSD50082.2020.9303081.
- [15] S. Radhimeenakshi, "Classification and prediction of heart disease risk using data mining techniques of Support Vector Machine and Artificial Neural Network," *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 3107-3111.

- [16] T. G. Keshavamurthy and M. N. Eshwarappa, "Review paper on denoising of ECG signal," *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 2017, pp. 1-4, doi: 10.1109/ICECCT.2017.8117941.
- [17] van Eck N. J., Waltman L. (2010) ' Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping', *Scientometrics* , 84/2: 523–38.

