5-10-2021

Breast Cancer Detection from Histopathology Images using Machine Learning Techniques: A Bibliometric Analysis

Shubhangi A. Joshi
Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, shubhangi.joshi.phd2020@sitpune.edu.in

Anupkumar M. Bongale Dr.
Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, ambongale@gmail.com

Arunkumar M. Bongale Dr.
Symbiosis Institute of Technology (SIT), Symbiosis International (Deemed) University, Pune, arunbongale1980@gmail.com

Breast Cancer Detection from Histopathology images using Machine Learning Techniques: A Bibliometric Analysis

Shubhangi A. Joshi[1], Anupkumar M. Bongale[2]*, Arunkumar Bongale[3]

[1] PhD Scholar,Symbiosis Institute of Technology, Symbiosis International (Deemed University), Lavale, Pune, Maharashtra, India
shubhangi.joshi.phd2020@sitpune.edu.in

[2] Department of Computer Science and Information Technology, Symbiosis Institute of Technology, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Lavale, Pune, Maharashtra, India
ambongale@gmail.com

[3] Department of Mechanical Engineering, Symbiosis Institute of Technology, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Lavale, Pune, Maharashtra, India
arunbongale1980@gmail.com

Abstract. Computer aided diagnosis has become upcoming area of research over past few years. With the advent of machine learning and especially deep learning techniques, the scenario of work flow management in healthcare sector is changing drastically. Artificial intelligence has shown potential in the field of breast cancer care. With datasets for machine learning frameworks getting eventually richer with time, we can definitely get newer insights in the field of breast cancer care. This will help in narrowing down the treatment range for patients and increasing patient survivability. The purpose of this study was to perform bibliometric analysis of the literature in the area of breast cancer detection using machine learning. Analysis was done for various elements like publication types, highly influential authors, most prominent journals, institutional affiliations, main keywords, etc. This analysis may direct future researchers by giving thorough quantitative evaluation of research documents in the field of breast cancer detection using machine learning.

Keywords: Bibliometric Analysis · Bibliometrics · Breast cancer detection · Deep learning · Machine learning · Histopathology

1 Introduction

Breast cancer is among the most common cancers found in women worldwide and it is the main cause of cancer related deaths. Early detection and timely treatment of breast cancer is very much important to improve life expectancy of patients. Machine learning and deep learning techniques have become popular tools for knowledge discovery in the field of healthcare. Despite of its several benefits, the mammography screening technique for breast cancer detection is associated with risk of false results [30]. Hence, an automated system for breast cancer detection and grade classification based on histopathology images can play a major role in computer-aided breast cancer diagnosis or prognosis. The increased need of accurate analysis of histopathological images for cancer detection has encouraged introduction of artificial intelligence systems in digital pathology [1, 22, 34, 47]. These systems can work as a digital assistant for oncologists and histopathologists. In breast cancer, cytological architecture of healthy cells is changed and these cells grow out of control resulting in formation of a carcinogenic mass called as tumor. There are two major types of breast cancers: Ductal carcinoma in Situ (DCIS) and Invasive carcinoma. Grade of breast cancer is an important factor in prognosis and it is also a representative of the aggressiveness of the tumor. Knowledge of type and grade of breast cancer is important for deciding line of treatment for the patient.

There are different diagnostic modalities used for breast cancer detection. Some of the common modalities are Magnetic resonance imaging (MRI), Breast ultrasound, Mammography, Positron emission tomography (PET), Fine needle aspiration or surgery to get tissue of suspected area (histopathological images), etc [18]. Analysis of medical data generated from these modalities is very time-consuming process. Even though all these modalities are widely used, cellular analysis of biopsy images seems to be the only way to detect presence of cancerous mass in the breast with confidence [19]. Due to criticality in the image analysis processes for breast cancer detection and grade classification, the experts are usually overburdened. This necessitates the automated workflow management system in the field of breast cancer care. The sample breast tissues aspired from suspected area of breast are usually stained with Hematoxylin and Eosin (H&E) stains. Eosin being acidic stain reacts with the basic part of breast tissue like cytoplasm and extracellular mass, staining them in pink color. And, due to its basic nature, hematoxylin stains the nuclei in purple or blue color. The acidic and basic properties of these stains give blue and pink colors to to the tissue structure. The H&E-stained images are true color RGB images having color depth of 24 bits. Histopathological image stained with H&E is shown in Figure 1.

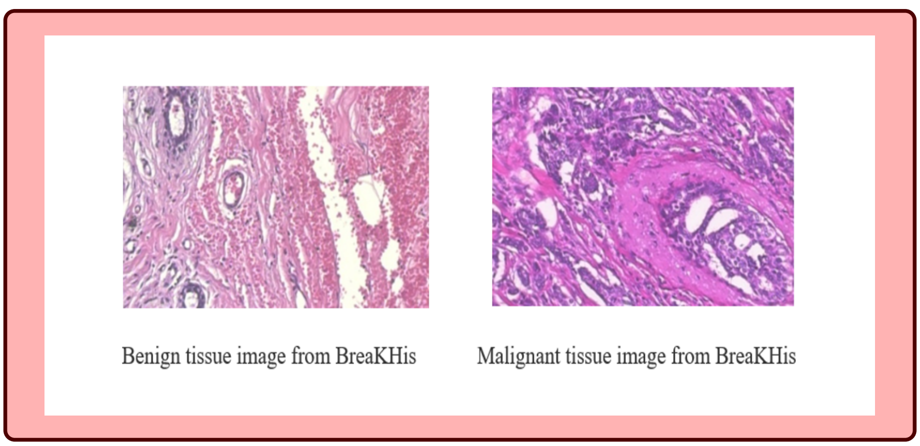

Fig. 1: Benign & malignant histopathology images(Source: BreaKHis dataset[2])

The analysis of H&E-stained histopathological images for cancer detection requires cell nuclei to be detected. There is lot of heterogeneity present in the cellular structure which complicates the task of nuclei detection. For nuclei detection and cancer type classification earlier work was focused on using traditional machine learning approaches such as weighted Naïve Bayes classifier [25],Linear discriminant analysis(LDA) classifier [8], Support vector machines [3, 33, 38, 40] etc. In one more research study, comparative analysis of classifiers for detecting breast cancer such as Naïve Bayes, Random Forests ,Rotation Forest was done [37]. But, most of the recent studies use transfer learning or deep learning approaches with the advent of new deep learning frameworks providing solutions to computational pathology problems like breast cancer detection. In one of the studies, a convolutional neural network which was spatially constrained was used for nuclei detection. Patch base classification was done for nuclei identification [39]. In one more study, stacked sparse autoencoder was used for locating individual nuclei in the given patch. The final SoftMax classifier was used for classification of image into either nucleus or non-nucleus image. Performance of deep learning structure against shallow learning structure was also compared [45]. Fully convolutional autoencoders were used for classification task in [28]. In one more research article focusing towards nuclei segmentation, U-net architecture was used with concave point detection algorithm. Overlapping and touching nuclei were also detected successfully [42].

In 2016 Spanhol and team [40] created BreaKHis dataset which was constructed from the histopathological images collected over the year of 2014 from PD Laboratory, Brazil. This dataset consists of total 7909 images of four different magnification factors such as 40X, 100X, 200X and 400X collected from 82 patients. These images are true color RGB images having color depth of 24 bits. Apart from providing a very good and benchmark dataset of histopathological images, this research also classified images into cancer and non-cancer binary classes.This dataset was used by many researchers for finding newer insights in breast cancer detection and classification task [9]. In the research conducted by Shallu et al. [31], the performance evaluation of transfer learning approach versus fully trained networks was observed. The three pre-trained network frameworks VGG16, VGG19 and ResNet50 were taken into consideration for experimentation. For transfer learning, these pretrained networks were used as feature generators and their output was fed to logistic regression classifier. The pre-trained model VGG16 which was fine-tuned, gave the best performance. In full trained network, the architecture of pre-trained model was used and weights were initialized randomly. It was demonstrated that performance of transfer learning (earlier trained on natural images but used for histopathological images) models was better than fully trained network even for limited dataset. In one more research work, four classes of breast cancer were identified using class structured deep convolutional neural network (CSDCNN). In CSDCNN, the feature learning was done automatically. End to end training was done for CSDCNN which learned semantic features and discriminative features in a bottom to top analysis manner. It took into consideration the feature space correlation for inter-class and intra-class features. It was found that distances of feature spaces from the samples of same class were significantly different from distances of feature spaces form the samples of different class. Classification was done by maximizing Euclidian distance of inter-class features. Validation and testing phases used real-world patient data. SoftMax was used as a final layer multi-class classifier [17].In one approach general classifier neural net was used for cancer detection from morphologically extracted features of breast tissue image. [21]

A novel method for breast carcinoma detection was proposed by A. Reddy and B. Soni [41]. Here support values concept was introduced for deep neural network. Algorithm for this model along with associated mathematical formula was devised. In this work changes in the mathematical equations of Sigmoid and Histo-Sigmoid functions were made. Histo-Sigmoid based method called as fuzzy-clustering was employed. In the research conducted by Dalal Bardou [6] two approaches for benign or malignant classification were considered. In first approach hand-crafted features were fed to SVM while in second approach CNN was used for classification task. Convolutional neural network approach was superior than approach using handcrafted features with SVM classifier. For first approach, features were extracted using local descriptors (DSIFT and SURF).

For second approach, BVLC Caffe CNN architecture with five convolutional layers was used. Many more research studies as in [2, 5, 7, 15, 23, 26, 35] used convolutional neural networks for classification of breast cancer. They used different methods for image preprocessing. Auto-classification of epithelial cells and stromal cells was done in [44]. Mitotic figures in histopathological slides were found using convolutional neural network in [4]. Tumor extent was quantified by deep learning approach on whole slide images in [11]. Many of the recent approaches included transfer learning frameworks which were initially trained on some other image based datasets [10, 12, 14, 36]. Generative adversarial networks were also used to discriminate the wrongly annotated patches in data augmentation methods as done in one of the studies [29]. Recently deep learning approaches were used for this classfication task as done in [13, 17, 20, 24, 43, 46]. Fusion of multiple deep learning models as an ensemble method for cancer detection was done in the studies [16, 27, 32].

In this paper a bibliometric analysis of research work done in the field of breast cancer detection using machine learning is discussed. Even though there are several imaging modalities for breast cancer detection, histopathological analysis of breast tissue serves as a gold standard for detecting breast cancer with confidence. And, hence research in the area of machine learning for developing automated diagnostic tool for breast cancer detection by making use of histopathological images has more clinical value. Breast cancer diagnostic research documents which were focusing on histopathological data are considered for this analytical study. For this analysis, research work available on Scopus database is referred.

The rest of the article is arranged as follows. Section 2 describes the keyword string used in search engines of various databases. The preliminary analyses of data collected from Scopus database viz. linguistic analysis, publication types, year-wise trend of publications are also presented in this section. Bibliometric information of the collected research documents is described in section 3. This section shows subject areas in which documents are published, geographical information regarding publications, important journals for the said area and the affiliating institutes where this research is conducted. Information regarding important keyword and citations is also presented in this section. Pictorial representation of information is given throughout the paper by using bibliometric analysis tools such as VOSviewer, ScienceScape and BiblioShiny(RStudio) environment. Deductions from this analytical study are presented in section 4. Future directions and scope for research in the area of breast cancer detection are discussed in section 5. And, finally paper is concluded in section 6.

2 Data Collection

Several data repositories like Scopus, Web of Science and IEEE Xplore were checked for availability of material on the aforementioned topic. The keywords used for forming search query are as given in Table 1

Table 1: Keywords used for generating search query String to Scopus Database (Source: Scopus DB accessed on 30th March 2021)

Primary Keywords	"breast cancer"
Supportive Keywords	detection OR classification
Supportive Keywords	"machine learning" OR "deep learning"
Supportive Keywords	histopahtology

The resultant search query string is as follows (TITLE-ABS-KEY (breast cancer) AND TITLE-ABS-KEY (detection OR classification) AND TITLE-ABS-KEY (("machine learning" OR "deep learning")) AND TITLE-ABS-KEY (histopathology))

This resultant search query string was applied to various databases on 30th March 2021. The number of articles obtained from selected databases is as given in Table 2

The most popular database with large number of highly cited and peer reviewed articles which provides data analytic tools is Scopus database. This database has a collection of documents from various domains such as engineer-

Table 2: Number of documents available from various databases

Sr.No.	Data Repository	No. of documents
1	Scopus	220
2	Web of Science	73
3	IEEE Xplore	63

ing, computer science, medicines, mathematics, physics, social sciences, etc. In this paper we have explored the analyzed the results for the research statement "Breast Cancer Detection from Histopathology images using Machine Learning Techniques" from Scopus database.

2.1 Basic Search Results from Scopus database

The Scopus database provided 220 documents in response to the search query string.The documents found on Scopus database are over the span of years from 2009 to 2021. Apart from 220 documents, there are **5 secondary documents** which are included in reference sections of Scopus documents, but not included in Scopus database. There are total **497 patents** filed on ideas related to this research topic which signifies potential capability of practical implementation of this research idea.428 patents are filed through United States Patent & Trademark Office. Linguistic analysis of documents from Scopus database is given in Table 3

Table 3: Publication language details
(Source: Scopus DB accessed on 30th March 2021)

Publication Language	Count
English	218
Chinese	1
Turkish	1

Total **218 documents** from Scopus database published in English language are considered for further analysis.

2.2 Publication Type Analysis

It is seen from publication type analysis that most of the documents are from journals and conferences indexed under Scopus. The details for the type of documents are shown in Table 4 and pie chart indicating various publication types is shown in Figure 2.

Table 4: Document type analysis
(Source: Scopus Database accessed on 30th March 2021)

Sr.No.	Publication type	Publication count	Percentage
1	Articles	131	60.09
2	Conference Papers	65	29.82
3	Review Papers	13	5.96
4	Conference Review	4	1.83
5	Book Chapter	3	1.38
6	Note	1	0.46
7	Short survey	1	0.46
	Total	218	100

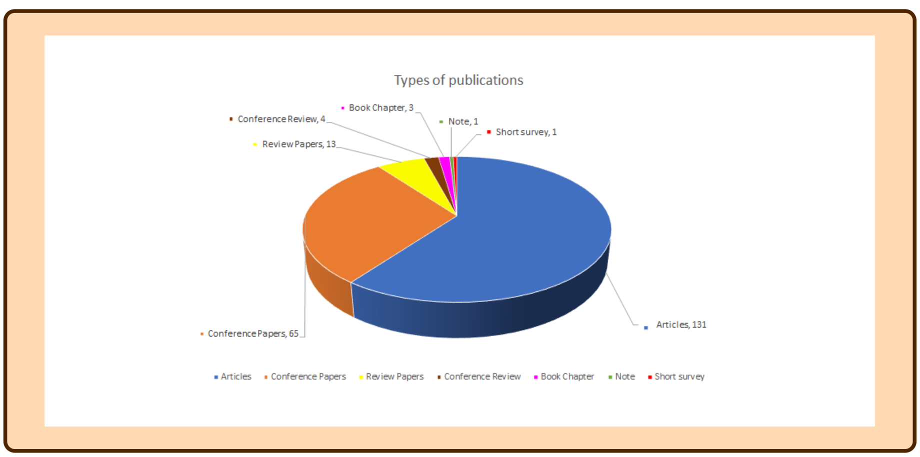

Fig. 2: Pie Chart for Publication Type Analysis
(Source: Scopus Database accessed on 30th March 2021)

2.3 Year-wise trend of Publications

The first publication regarding aforementioned topic is found in the year of 2009.It is seen that very few documents were published till the year of 2015.After the advancements in digital pathology techniques and parallel technological growth in the field of machine learning and deep learning evolved large surge of research in the field of breast cancer detection using histopathology.As evident from Table 5 and shown in Figure 3, highest number of research documents were published in the year of 2019 and 2020.It means that the researchers who are currently working in this field are in the most flourishing time frame of this domain as favourable resources must be available for research with lot of scope for improvements in earlier work done.

Table 5: Year-wise trend of publications
(Source: Scopus Database accessed on 30th March 2021)

Year	**Publication trend**	**Year**	**Publication trend**
2021	23	2014	3
2020	79	2013	1
2019	55	2012	1
2018	26	2011	1
2017	12	2010	2
2016	12	2009	1
2015	2		

3 Bibliometric Analysis

From histopathological images extracted from suspected breast area, pathologists can find out type of breast cancer (DCIS or Invasive) as well as grade of breast cancer. In order to relieve burden on pathologist in the areas where medical task force is scares, such analysis performed by automated diagnostic technique will prove very beneficial. Breast cancer detection process deals with histological analysis to find out the presence of cancer. Breast cancer classification allows the type of breast cancer to be found out so that suitable treatment

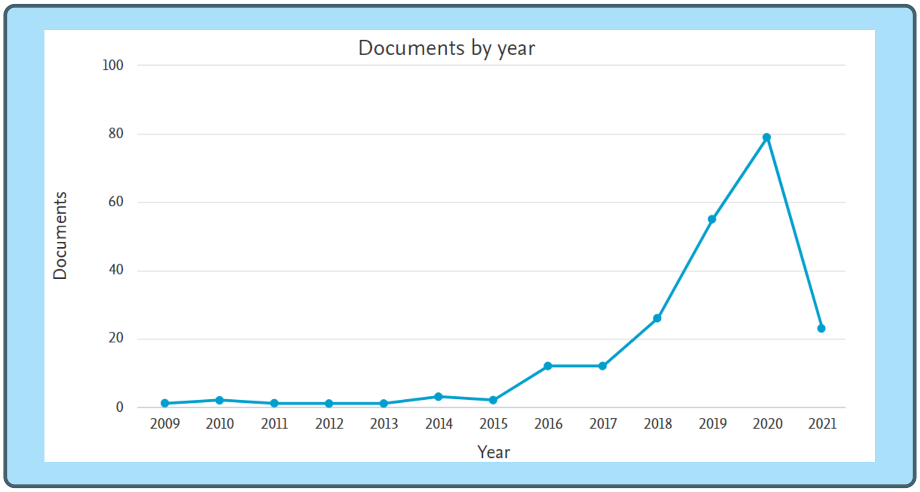

Fig. 3: Year-wise trend of publications
(Source: Scopus Database accessed on 30th March 2021)

can be given to the patients. Many researchers have contributed to the development of computer aided techniques for automatic detection and classification of breast cancer. Most of the work revolves around using latest techniques of machine learning and deep learning under the broad umbrella of artificial intelligence.

3.1 Analysis of Subject Areas

The main subject area in which this research is conducted is found to be the area of Computer Science. This is very obvious as artificial intelligence techniques are one of the key areas in computer science. The next dominant area of research is medicines followed by engineering stream.

The pie-chart shown in Figure 4 clearly indicates that large portion of this work is getting done in the field of computer science.

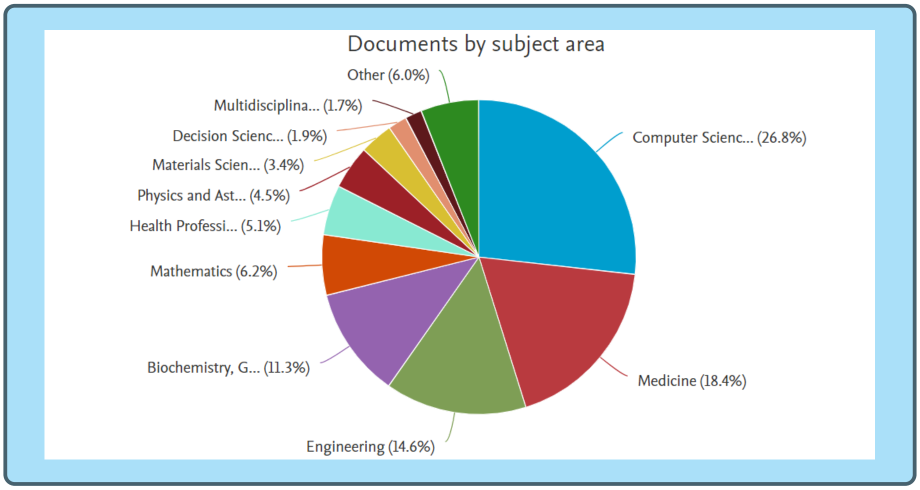

Fig. 4: Subject areas under which research is conducted
(Source: Scopus Database accessed on 30th March 2021)

3.2 Geographical Analysis

This research in the field of breast cancer care is being carried out in worldwide manner. Many of the developing and developed countries have thrust towards research in the field of healthcare.

The above graph in Figure 5 depicts that Indian researchers have dominated the worldwide research community by standing at first position. The next highest researching country is United States. This clearly gives indication of good resourceful environment for Indian researchers in the domain of breast cancer analysis.

The country-based citation analysis is carried out in VOSviewer software by selecting countries which have minimum 5 number of publications with minimum 5 citations.

Total 15 countries matched the threshold. For these 15 countries, total strength of citation links with other countries is calculated and represented in following network visualization as shown in Figure 6 The country-based network map for maximum co-authorship is generated in VOSviewer and is given in the Figure

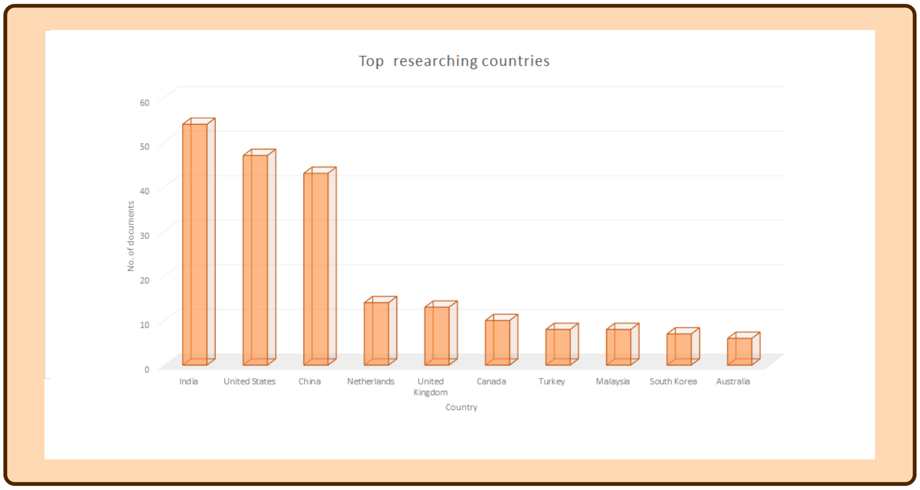

Fig. 5: Top Ten Countries in publication
(Source: Scopus Database accessed on 30th March 2021)

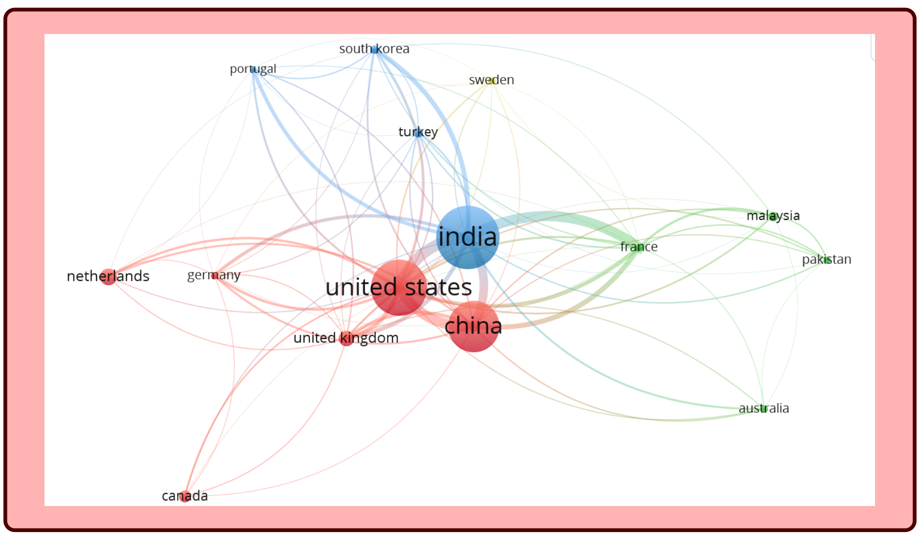

Fig. 6: Network Visualization of top Fifteen Countries in publication with maximum citations
(Source: VOSviewer)

7 below. Minimum number of documents of a country are limited to 5. Co-authorship among countries is shown in the form of links of network map.

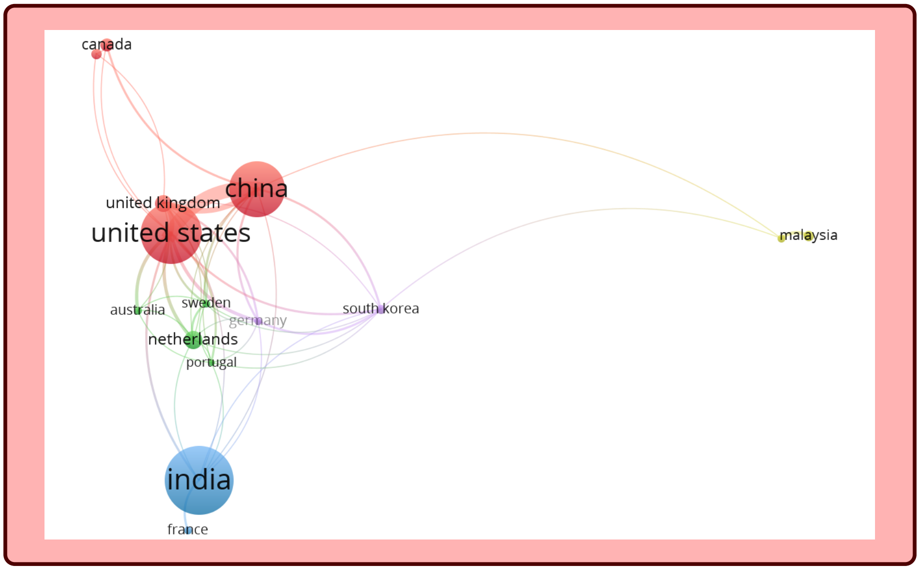

Fig. 7: Network Connection diagram of researching countries with co-authorship (Source: VOSviewer)

From above analysis, it is seen that India, United States and China are the nations showing very strong links of co-authorship among countries.These countries also exhibit maximum citations for their publications.

Different countries also encourage collaborative work for breast cancer care using artificial intelligence.The collaboration links between different countries can be seen from Figure 8.This collaboration map is built using Biblioshiny tool supported by RStudio.

3.3 Analysis of sources

Most influential sources of publications are given in the Figure 9. This information is extracted from Biblioshiny tool supported by RStudio.These all journals are highly reputed with very good impact factor. All these top journals are in the field of computer science, engineering , medicines,etc

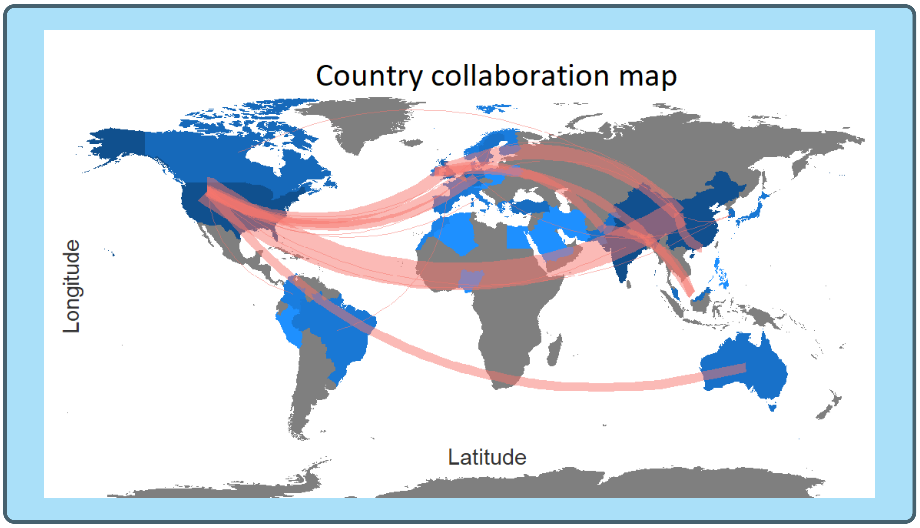

Fig. 8: Collaboration world Map
(Source: Biblioshiny by RStudio)

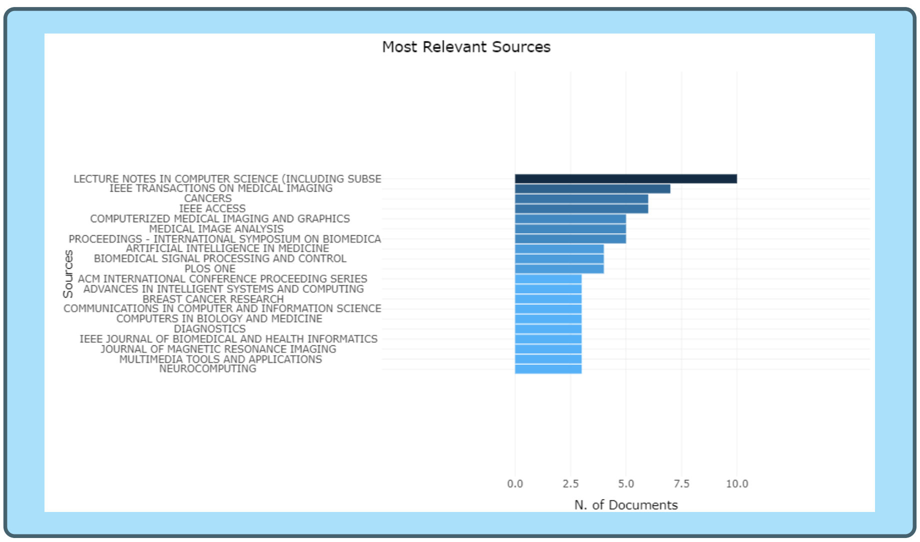

Fig. 9: Most relevant sources of publications
(Source: Biblioshiny by RStudio)

3.4 Affiliation Analysis

The affiliations analysis indicates the main organizations that are engaged in the research on breast cancer care using machine learning. This information is tabulated in the Table 6 The graph for the affiliations is shown in Figure 10 below. The analysis shows Radboud university medical center, Netherlands has published major documents followed by Case Western Medical center, United States.

Table 6: Affiliation of organizations for the research

Sr.No.	**Organization**	**Documents**
1	Radboud University Nijmegen Medical Centre, Netherlands	6
2	Case Western Reserve University, United States	5
3	University Hospitals Case Medical Centre, Ohio, United States	5
4	Huazhong University of Science and Technology, China	5
5	Shenzhen University, China	4
6	University of Warwick, England	4
7	University of California, Los Angeles, United States	4
8	Technische Universiteit Eindhoven	4
9	Nanjing University of Information Science & Technology, China	4
10	Sorbonne Universite, France	3

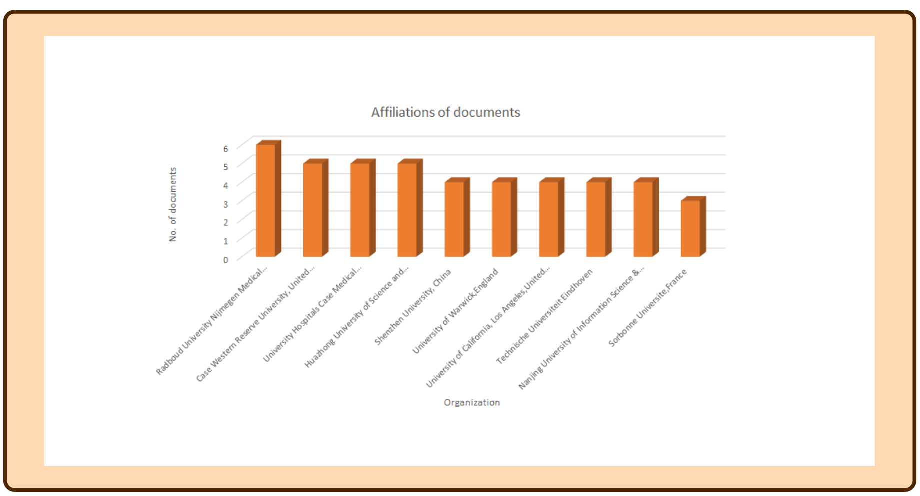

Fig. 10: Affiliation graph for various documents
(Source: Scopus DB accessed on 30th March 2021)

3.5 Keyword Analysis

Different keywords used in this research study are breast cancer, detection, classification, machine learning, deep learning, artificial intelligence.Table 7 below shows the frequency of occurrence of keywords used in the articles published in this domain. Frequency of co-occurrence of keywords was limited to 15 while forming network diagram in VOS Viewer.

Table 7: Occurrence of top keywords

Sr.No.	**Keyword**	**Count**
1	Breast cancer	165
2	Histopathology	124
3	Humans	112
4	Article	94
5	Deep learning	140
6	Medical imaging	115
7	Diseases in Female	183
8	Machine learning	78
9	Breast neoplasm	57
10	Breast Tumor	55

The keyword-based network map is shown in Figure 11. Minimum strength of linking lines is kept as 15.

Figure 12 shows density visualization of keywords. Highest density is indicated by yellow-coloured patches.

It is seen that, most dominant keywords in the publications from Scopus database are breast cancer, histopathology, deep learning, medical imaging, machine learning, etc. Many of the important keywords which occur in abstract section of the documents is shown in the keyword treemap shown in Figure13. This treemap is generated by using biblioshiny tool supported by RStudio. Figure 14 indicates the evolution of important keywords in the research area over the span of 2009 to 2021.Most influential and research oriented keywords have shown substantial growth in the time frame of 2018 to 2021.This pattern for

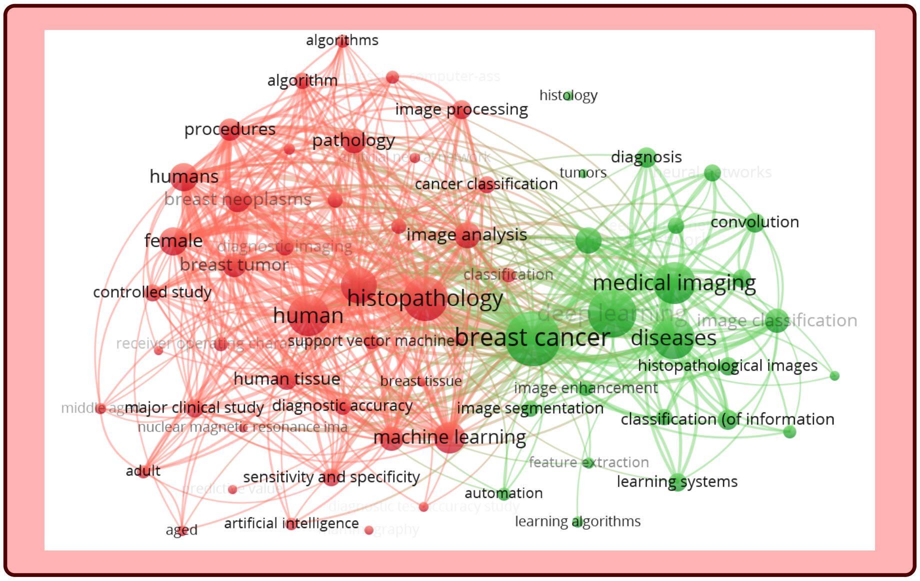

Fig. 11: Network map for keyword co-occurrence
(Source: VOSviewer)

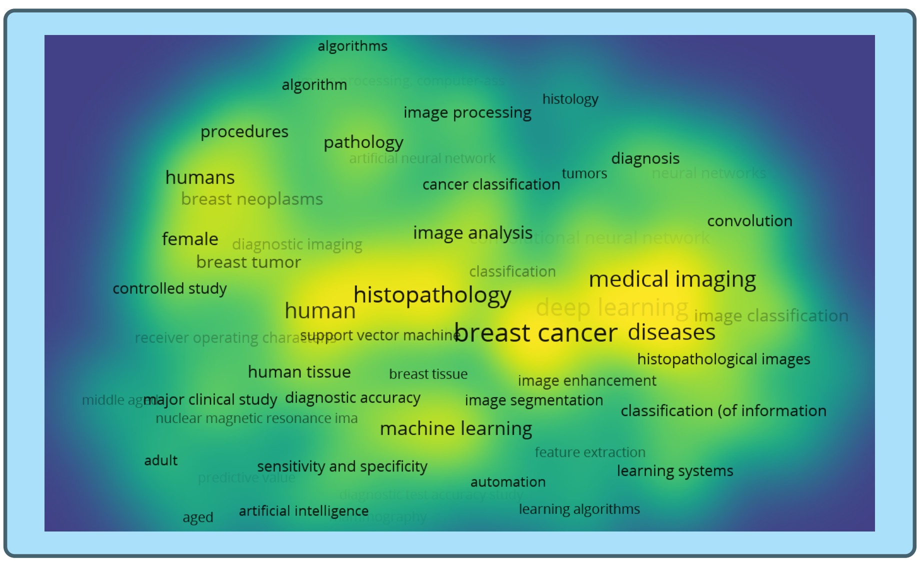

Fig. 12: Density diagram for keyword co-occurrence
(Source: VOSviewer)

Fig. 13: Treemap for keyword occurrence in the abstract section of documents (Source: Biblioshiny by RStudio)

keyword evolution is extracted from Sciencescape website which is widely used for scientometric analysis.

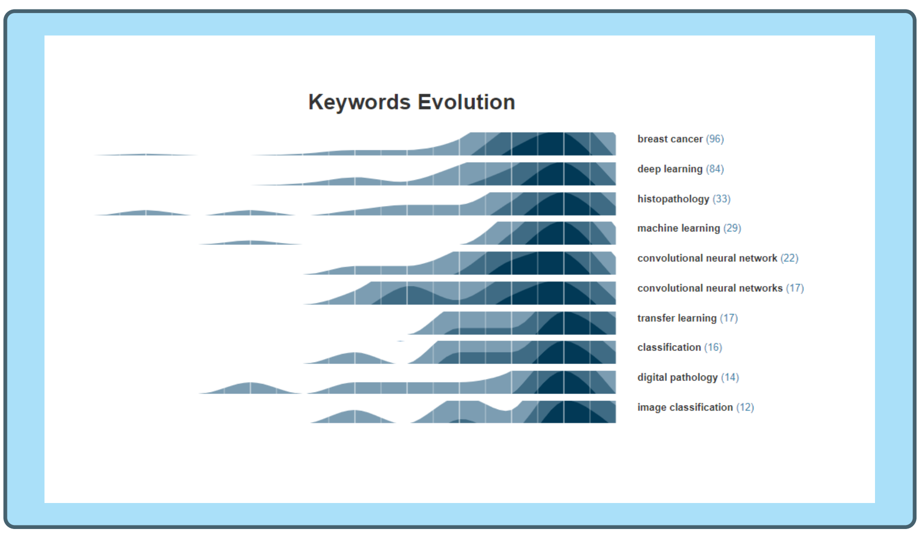

Fig. 14: Evolution of keywords over time (Source: ScienceScape)

3.6 Citation Analysis

Top 10 highly cited papers having total citations more than 100 are shown in the Table 8.It can be seen that greatest number of citations are done in the span of year 2018 to 2021.This indicates current importance of this topic of research.

Table 8: Top ten highly cited research documents available on Scopus database

Document Title	**Author**	**Year of Publication**	**Year-wise Citations**						**Total Citations**
			<2017	2017	2018	2019	2020	2021	
Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images	Korsuk Sirinukunwattana, Shan E Ahmed Raza	2016	5	70	91	156	138	33	493
Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images	Jun Xu, Lei Xiang, Qingshan Liu, Hannah Gilmore	2016	10	60	85	129	92	23	399
Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review—Current Status and Future Potential	Humayun Irshad, Antoine Veillard	2014	67	72	56	77	69	18	359
A Dataset for Breast Cancer Histopathological Image Classification	Fabio A. Spanhol , Luiz S. Oliveira	2016	3	22	51	106	102	41	322

Document Title	**Author**	**Year of Publication**	**Year-wise Citations**						**Total Citations**
			<2017	2017	2018	2019	2020	2021	
Classification of breast cancer histology images using Convolutional Neural Networks	Teresa Araujo, Guilherme Aresta	2017	0	3	52	100	116	39	310
AggNet: Deep Learning From Crowds for Mitosis Detection in Breast Cancer Histology Images	Shadi Albarqouni, Christoph Baur	2016	7	31	53	67	62	27	247
International evaluation of an AI system for breast cancer screening	Scott Mayer McKinney,Marcin Sieniek	2020	0	0	0	0	174	71	245
A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images	Jun Xu a,n , Xiaofei Luo a , Guanhao Wang a , Hannah Gilmore	2016	5	43	56	69	62	10	245
Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI	Nathaniel M. Braman, Maryam Etesami	2017	0	1	24	59	58	16	158
Deep Learning for Magnification Independent Breast Cancer Histopathology Image Classification	Neslihan Bayramoglu,Juho Kannala	2016	0	10	18	37	40	20	128

The maximum citation count obtained for nuclei classification paper is 490. The highly influential authors with highest number of citations are given in the Table 9

Table 9: Most influential Authors

Sr.No.	Name of Author	No. of documents	Total citations
1	H. Gilmore	5	839
2	A. Madabhushi	8	895
3	J. Xu	4	702
4	D. Racoceanu	4	432
5	G. Murtaza	3	6
6	G. Raza	3	6
7	I. Shuib	3	6
8	H. Chen	3	39
9	C. Li	5	118
10	Q. Wang	3	25

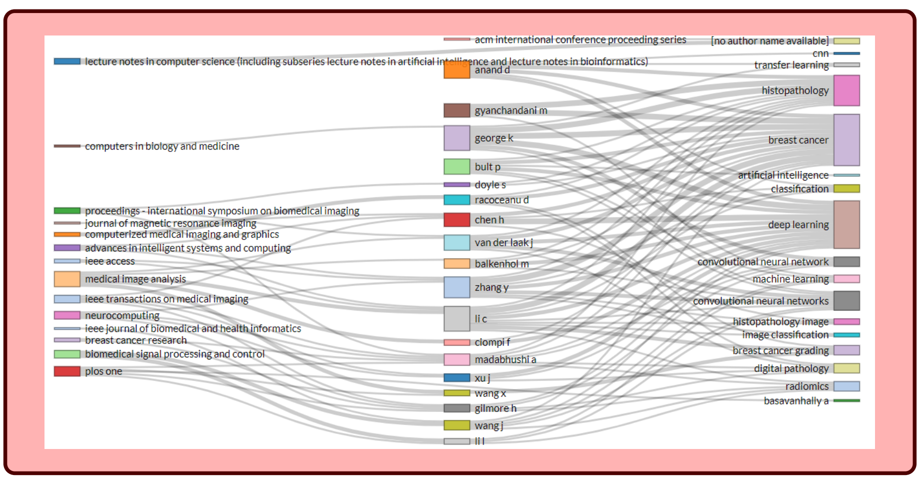

Fig. 15: Three fields plot representing most influential journals, authors and keywords
(Source: Biblioshiny by RStudio)

Three fields plot which was extracted from Biblioshiny software supported by RStudio, gives correlation of authors with high impact, keywords and associated journals. It is given in Figure 15

4 Deductions from Bibliometric Analysis

Technological advances make it possible to use machine learning for computational pathology. Insights given by computer aided diagnostic machines are very useful for correct diagnosis and prognosis of any disease. Computational pathology using deep learning is very niche idea which is getting used worldwide. It could not have been possible 10 years ago. In breast cancer care, machine learning and deep learning are involved in providing error free solutions to exiting pathological problems. Lot of work is being carried out for malignancy detection, classification and grading of cancer, finding molecular drivers of cancer cells, etc. Most of the research in the domain of breast cancer care using machine learning has been accelerated with the advent of deep learning technologies i.e., from the year 2017-2018. There is lot of research being carried out in this domain in developing and developed countries of the world. Maximum number of publications are found in English language. Keyword analysis implies implementation of breast cancer detection algorithms by using machine learning and deep learning techniques under the broader umbrella of artificial intelligence.

Due to digitalization and computer aided automation, pathological data organization has become easy. This makes work of lab technicians/histopathologists less stressful.

5 Scope for future research

Due to various existing challenges, in-depth research studies for cancer diagnosis and prognosis are still not completely unlocked. Analysis based on multiple imaging modalities is possible as a part of future advancement in latest studies. As datasets keep becoming richer and richer, these large volumes of data will make more transitions in current automation models used in digital pathology related to cancer care. After 10 years, with the availability of richer volumes of data, more and more companies will participate in innovating newer tools for computational pathology and it is very likely that pathologists will start using these Artificial Intelligence tools as a kind of digital assistance to them. This field of computational pathology has a potential to change complete business model of pathology in future. Computer aided diagnosis with digital pathology will make healthcare work-flow very efficient.

6 Conclusion

In this bibliometric paper, quantitative analysis of research trends in the domain of breast cancer detection using machine learning techniques is done. Research documents using histopathological modality for the analysis were focused. All documents used for analysis were collected from Scopus database. Information related to most influential authors, institutes where these types of researches are done ,most important journals , highly cited articles,etc is analysed in this paper.Similarly keyword analysis , authorship co-occurrence analysis are also evaluated. Overall quantitative evaluation indicates good scope for further research in the field of breast cancer care by using machine learning or deep learning approaches.

References

1. Acs, B., Rantalainen, M., Hartman, J.: Artificial intelligence as the next step towards precision pathology. Journal of internal medicine **288**(1), 62–81 (2020)
2. Adeshina, S.A., Adedigba, A.P., Adeniyi, A.A., Aibinu, A.M.: Breast cancer histopathology image classification with deep convolutional neural networks. In: 2018 14th international conference on electronics computer and computation (ICECCO). pp. 206–212. IEEE (2018)
3. Akinnuwesi, B.A., Macaulay, B.O., Aribisala, B.S.: Breast cancer risk assessment and early diagnosis using principal component analysis and support vector machine techniques. Informatics in Medicine Unlocked **21**, 100459 (2020)

4. Albarqouni, S., Baur, C., Achilles, F., Belagiannis, V., Demirci, S., Navab, N.: Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images. IEEE transactions on medical imaging **35**(5), 1313–1321 (2016)
5. Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A., Campilho, A.: Classification of breast cancer histology images using convolutional neural networks. PloS one **12**(6), e0177544 (2017)
6. Bardou, D., Zhang, K., Ahmad, S.M.: Classification of breast cancer based on histology images using convolutional neural networks. Ieee Access **6**, 24680–24693 (2018)
7. Bayramoglu, N., Kannala, J., Heikkilä, J.: Deep learning for magnification independent breast cancer histopathology image classification. In: 2016 23rd International conference on pattern recognition (ICPR). pp. 2440–2445. IEEE (2016)
8. Belsare, A., Mushrif, M., Pangarkar, M., Meshram, N.: Classification of breast cancer histopathology images using texture feature analysis. In: TENCON 2015-2015 IEEE Region 10 Conference. pp. 1–5. IEEE (2015)
9. Benhammou, Y., Achchab, B., Herrera, F., Tabik, S.: Breakhis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights. Neurocomputing **375**, 9–24 (2020)
10. Chang, J., Yu, J., Han, T., Chang, H.j., Park, E.: A method for classifying medical images using transfer learning: A pilot study on histopathology of breast cancer. In: 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom). pp. 1–4. IEEE (2017)
11. Cruz-Roa, A., Gilmore, H., Basavanhally, A., Feldman, M., Ganesan, S., Shih, N., et al.: Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent. sci rep. 2017; 7: 46450
12. Deniz, E., Şengür, A., Kadiroğlu, Z., Guo, Y., Bajaj, V., Budak, Ü.: Transfer learning based histopathologic image classification for breast cancer detection. Health information science and systems **6**(1), 1–7 (2018)
13. Feng, Y., Zhang, L., Mo, J.: Deep manifold preserving autoencoder for classifying breast cancer histopathological images. IEEE/ACM transactions on computational biology and bioinformatics **17**(1), 91–101 (2018)

14. Golatkar, A., Anand, D., Sethi, A.: Classification of breast cancer histology using deep learning. In: International Conference Image Analysis and Recognition. pp. 837–844. Springer (2018)
15. Gour, M., Jain, S., Sunil Kumar, T.: Residual learning based cnn for breast cancer histopathological image classification. International Journal of Imaging Systems and Technology **30**(3), 621–635 (2020)
16. Hameed, Z., Zahia, S., Garcia-Zapirain, B., Javier Aguirre, J., María Vanegas, A.: Breast cancer histopathology image classification using an ensemble of deep learning models. Sensors **20**(16), 4373 (2020)
17. Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., Li, S.: Breast cancer multi-classification from histopathological images with structured deep learning model. Scientific reports **7**(1), 1–10 (2017)
18. Houssein, E.H., Emam, M.M., Ali, A.A., Suganthan, P.N.: Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. Expert Systems with Applications p. 114161 (2020)
19. Irshad, H., Veillard, A., Roux, L., Racoceanu, D.: Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. IEEE reviews in biomedical engineering **7**, 97–114 (2013)
20. Islam, M.M., Huang, S., Ajwad, R., Chi, C., Wang, Y., Hu, P.: An integrative deep learning framework for classifying molecular subtypes of breast cancer. Computational and structural biotechnology journal **18**, 2185–2199 (2020)
21. Jain, A., Atey, S., Vinayak, S., Srivastava, V.: Cancerous cell detection using histopathological image analysis. Int J Innov Res Comput Commun Engng **2**(12), 7419–26 (2014)
22. Janowczyk, A., Madabhushi, A.: Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. Journal of pathology informatics **7** (2016)
23. Jiang, Y., Chen, L., Zhang, H., Xiao, X.: Breast cancer histopathological image classification using convolutional neural networks with small se-resnet module. PloS one **14**(3), e0214587 (2019)
24. Karthiga, R., Narashimhan, K.: Deep convolutional neural network for computer-aided detection of breast cancer using histopathology images. In: Journal of Physics: Conference Series. vol. 1767, p. 012042. IOP Publishing (2021)

25. Kharya, S., Soni, S.: Weighted naive bayes classifier: a predictive model for breast cancer detection. International Journal of Computer Applications **133**(9), 32–37 (2016)
26. Kumar, K., Rao, A.C.S.: Breast cancer classification of image using convolutional neural network. In: 2018 4th International Conference on Recent Advances in Information Technology (RAIT). pp. 1–6. IEEE (2018)
27. Lee, S., Amgad, M., Masoud, M., Subramanian, R., Gutman, D., Cooper, L.: An ensemble-based active learning for breast cancer classification. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 2549–2553. IEEE (2019)
28. Li, X., Radulovic, M., Kanjer, K., Plataniotis, K.N.: Discriminative pattern mining for breast cancer histopathology image classification via fully convolutional autoencoder. IEEE Access **7**, 36433–36445 (2019)
29. Man, R., Yang, P., Xu, B.: Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks. IEEE Access **8**, 155362–155377 (2020)
30. McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.S., Darzi, A., et al.: International evaluation of an ai system for breast cancer screening. Nature **577**(7788), 89–94 (2020)
31. Mehra, R., et al.: Breast cancer histology images classification: Training from scratch or transfer learning? ICT Express **4**(4), 247–254 (2018)
32. Murtaza, G., Shuib, L., Wahab, A.W.A., Mujtaba, G., Raza, G.: Ensembled deep convolution neural network-based breast cancer classification with misclassification reduction algorithms. Multimedia Tools and Applications **79**(25), 18447–18479 (2020)
33. Obaid, O.I., Mohammed, M.A., Ghani, M., Mostafa, A., Taha, F.: Evaluating the performance of machine learning techniques in the classification of wisconsin breast cancer. International Journal of Engineering & Technology **7**(4.36), 160–166 (2018)
34. Robertson, S., Azizpour, H., Smith, K., Hartman, J.: Digital image analysis in breast pathology—from image processing techniques to artificial intelligence. Translational Research **194**, 19–35 (2018)
35. Saxena, S., Shukla, S., Gyanchandani, M.: Pre-trained convolutional neural networks as feature extractors for diagnosis of breast cancer using histopathology. International Journal of Imaging Systems and Technology **30**(3), 577–591 (2020)

36. Sebai, M., Wang, T., Al-Fadhli, S.A.: Partmitosis: A partially supervised deep learning framework for mitosis detection in breast cancer histopathology images. IEEE Access **8**, 45133–45147 (2020)
37. Shukla, K., Tiwari, A., Sharma, S., et al.: Classification of histopathological images of breast cancerous and non cancerous cells based on morphological features. Biomedical and Pharmacology Journal **10**(1), 353–366 (2017)
38. Singh, S., Kumar, R.: Histopathological image analysis for breast cancer detection using cubic svm. In: 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN). pp. 498–503. IEEE (2020)
39. Sirinukunwattana, K., Raza, S.E.A., Tsang, Y.W., Snead, D.R., Cree, I.A., Rajpoot, N.M.: Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE transactions on medical imaging **35**(5), 1196–1206 (2016)
40. Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L.: A dataset for breast cancer histopathological image classification. Ieee transactions on biomedical engineering **63**(7), 1455–1462 (2015)
41. Vaka, A., Soni, B.: Srk,". Breast cancer detection by leveraging Machine Learning," ICT Express (2020)
42. Wan, T., Zhao, L., Feng, H., Li, D., Tong, C., Qin, Z.: Robust nuclei segmentation in histopathology using asppu-net and boundary refinement. Neurocomputing **408**, 144–156 (2020)
43. Xie, J., Liu, R., Luttrell IV, J., Zhang, C.: Deep learning based analysis of histopathological images of breast cancer. Frontiers in genetics **10**, 80 (2019)
44. Xu, J., Luo, X., Wang, G., Gilmore, H., Madabhushi, A.: A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. Neurocomputing **191**, 214–223 (2016)
45. Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., Madabhushi, A.: Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images. IEEE transactions on medical imaging **35**(1), 119–130 (2015)
46. Yao, H., Zhang, X., Zhou, X., Liu, S.: Parallel structure deep neural network using cnn and rnn with an attention mechanism for breast cancer histology image classification. Cancers **11**(12), 1901 (2019)
47. Yari, Y., Nguyen, H., Nguyen, T.V.: Accuracy improvement in binary and multi-class classification of breast histopathology images. In: 2020 IEEE Eighth Interna-

tional Conference on Communications and Electronics (ICCE). pp. 376–381. IEEE (2021)