A Meta-Analysis of Evolution of Deep Learning Research in Medical Image Analysis

David ZENG
College of BIS, Dakota State University
Madison, SD 57042, United States

Cherie NOTEBOOM
College of BIS, Dakota State University
Madison, SD 57042, United States

Kruttika SUTRAVE
College of BIS, Dakota State University
Madison, SD 57042, United States

Rajesh GODASU
College of BIS, Dakota State University
Madison, SD 57042, United States

ABSTRACT [1]

With a text mining and bibliometrics approach, we review the literature on the evolution of deep learning in medical image literature from 2012 – 2020 to understand the current state of the research and to identify the major research themes in image analysis to answer our research questions: RQ1: What are the learning modes that are evident in the literature? RQ2: What are the emerging learning modes in the literature? RQ3: What are the major themes in medical imaging literature? The analysis of 8704 resulting from a data collection process from peer-reviewed databases, our analysis discovered the six major themes of image segmentation studies, studies with image classification, evaluation procedures such as sensitivity and specificity, optical coherence tomography studies, MRI imaging studies, and Chest imaging studies. Additionally, we assessed the number of articles published each year, the frequent keywords, the author networks, the trending topics, and connections to other topics. We discovered that segmenting and classifying the images are the most common tasks. Transfer learning is the most researched area and cancer is the highly targeted disease and Covid-19 is the most recent research trend.

Keywords: Deep Learning, Medical Image Analysis, Text Mining, VosViewer.

[1] The authors would like to express gratefulness to Kaushik Nagarajan, for his comprehensive and detailed peer-editing of this document.

1. INTRODUCTION

AI has a long history dating back to 1956 when the term Artificial Intelligence was first coined [1]. Algorithms called artificial neurons arranged into networks have led to programs that can learn. The Perceptron is one of the simplest neural network architectures, invented in 1957 [2]. However, only during the 1990s, the abilities of these neural networks were exploited with the introduction of Deep Learning (DL) i.e. development of multi-layered neural networks [3]. In 2012, a breakthrough in computer vision in ImageNet Large-Scale Visual Recognition Challenge [4] has led to the advancement of DL and the expansion of AI techniques in the healthcare domain [1].

Deep Transfer Learning (DTL) a sub-field of AI, has gained importance in the medical field as an effective tool to perform medical image analysis including classification, localization, detection, segmentation, and registration [5]. The crux of DTL is to apply knowledge obtained from models trained on large data sets to different but related tasks in which data sets may be limited. While Deep TL systems enable us to leverage knowledge from previously trained models to a different domain, it is a novel and challenging, yet promising field for research. The potential exists for DL/TL research discoveries to predict medical outcomes using images by enhancing the model effectiveness of solving the new problems when there exists divergence from the source to the target domains [6], [7].

DTL, however, possesses several potential disadvantages. Accurately labeled data are required for both source and target domains, but obtaining large amounts of labeled data is considered a fundamental challenge in medical imaging [8]. Hierarchical representations learned from pre-trained models (under the source domain) may not be very insightful in solving the new task when the domains are very different because the convolutional layers near input contain generic features and layers near the output contain features specific to the target problem [9], [10]. To mitigate the limitations of traditional DTL, in the latest studies [11]–[13] different learning paradigms are emerging - Reinforcement Learning (RL) and Federated Learning (FL). In RL, the model learns from interactions between its actions and environment contrary to learning directly from feedback signals calculated based on the differences between the target (ground truth) and prediction. FL enables multiple end devices to collaboratively learn a shared predictive model while keeping all the training data on the device, avoiding the process of aggregating massive data from every network edge to a remote cloud server and performing learning tasks at that remote server [14]–[16].

Deep Convolutional Neural Networks (DCNNs) and the accessibility of efficient and powerful graphical processing units (GPUs) have made it possible to significantly improve the performance of medical image analysis [17], [18]. DCNNs are the current state-of-the-art technique in medical image analysis, performing better than radiologists in some cases [19]. DCNN architectures require large, annotated datasets. Overfitting and lack of generalization are common issues encountered with standard CNNs when the training data provided are insufficient [20]. GANs are emerging as an effective method for increasing and improving training datasets by generating realistic artificial images [21].

Utilizing DL techniques in Medical image analysis is a fast-growing research area [5] with several studies being published every day in different journals. In today's world, peer-reviewed journals track current innovations and developments in any research field, submitting the work in a journal that is popular with a high impact factor leads to an increasing number of citations [22]. Measuring the research outcomes from these journals to gain insights into the evolution of any field is known as bibliometrics or scientometric analysis [23]. Our study combines the methods of text-mining and bibliometrics to assess the evolution of medical image analysis using DL techniques with the use of tools namely NLP and VOS-viewer [24].

VOS-viewer is a tool built for analyzing a large number of research articles with its clustering technique, the VOS (Visualization of Similarities) clustering method. It provides graphs that contain the mapping between the key topics to learn the relationship strengths among the topics with help of attributes such as "occurrences" and 'total link strength". Web of Science (WOS) is the preferred online library for VOS-viewer analysis [22], [25]. WOS tracks the most recent developments in all major research areas. Our study uses four different web-based libraries along with WOS to build a customized library with a comprehensive set of articles in medical image analysis. The complete search strategy is explained in Section 3.

The contributions of this Literature-Analysis are the following. First, to our knowledge, this is the first effort to explore the evolution of DL in medical image literature using the text-mining and bibliometrics approach. Second, we attempt to discover the issues, challenges, and opportunities to guide our future research projects. Third, while we define this proposal as a Literature-analysis work only, we will extend the findings to propose a multi-task and multi-mode learning framework for medical imaging as our next research effort that includes extending the research papers on different learning modes by utilizing a formal SLR methodology. The specific research questions of this study are: 1: What are the learning modes that are evident in the literature? 2: What are the emerging learning modes in the literature? 3: What are the major themes in medical imaging literature?

2. BACKGROUND

Convolutional Neural Networks

CNNs are a class of deep neural networks, widely utilized for image analysis. They contain three different types of layers (a) Convolutional layers, (b) Pooling layers, and (c) Fully connected layers [26]. Convolutional layers are the crux of the architecture consisting of moving filters. A convolution operation is sliding a filter across the width and height of the input image area (receptive field) to multiply the elements of the corresponding receptive field and the filter to produce a feature map [26].

Transfer Learning

A person who can play tennis can use the knowledge to learn to play baseball. Similarly, in the machine learning context, TL refers to a scenario where the features learned in one task are leveraged to improve the classification accuracy in another task. For instance, in task T1, a Deep CNN gaining insights from visual features of dog images will be able to use this knowledge to classify the images of cats in task T2. TL is proved to reduce the necessity of large amounts of datasets to some extent. This idea of learning knowledge from one task and applying it to a specific target task has led numerous studies recently to use pre-trained CNNs on ImageNet (natural images) [27] dataset for various image recognition tasks, especially in Medical Image classification [7], [26], [28]. TL can be implemented in two methods – feature extraction and fine-tuning.

Feature extraction is said to be performed when a CNN network learns features from Task 1, then uses the same base network of convolution and pooling layers by replacing the fully connected layer with Task 2 classifier. During this operation, the weights of the convolutional base are frozen and only a new classifier is trained for the classification of task 2. In fine-tuning, the weights of the convolutional base are slightly altered to match the task 2 problem after training the network on Task 1 data. Unlike feature extraction, fine-tuning requires unfreezing some top-level layers of convolution base and a new classifier is added. While performing this type of TL, the unfrozen layers of CNN are trained along with the fully connected layer when passed through new data [6].

Generative Adversarial Networks

GANs are generative models with an adversarial process that were first empirically demonstrated by Ian Goodfellow [29] in 2014. They are used to create new examples of data that are similar to training data. For instance, they can generate images of human faces based on examples from training data, although these faces do not belong to any real person. GANs consist of two components – the generator model and the discriminator model. The function of the generator is to create realistic artificial images and that of the discriminator is to classify whether an image is real or artificial. The generator model takes noise (Gaussian or standardized distribution) as input and learns to generate synthetic images whereas, the discriminator model receives both the generated and original images, it learns to distinguish between the two - real or fake images. Both the networks try to optimize a different and opposing objective function.

Image Segmentation

Image segmentation is a vital task for the quantitative analysis of medical images. It involves compartmentalizing images into multiple regions with similar properties [30]. Image segmentation is of two different types – semantic segmentation (grouping of pixels with semantic labels) or instance segmentation (grouping objects) [31]. Semantic segmentation involves labeling each pixel in the image. Instance segmentation is a further extension of semantic segmentation as it distinguishes each object of interest. In medical imaging, segmentation is used to study the anatomical structure and locate a region of interest (ROI) such as lesions and tumors. Segmentation is performed on various medical image modalities such as MRI, CT, X-ray, OCT, and PET. In the past, several researchers have proposed various segmentation techniques using deep learning methods. For instance, in [32], the authors use a supervised method using fully connected CNN along with transfer learning to segment retinal vessels. Authors of [33], present a method for brain masking using the U-net model. Another study [34] proposed a DL method for pericardium segmentation using CT images.

Image Classification

Medical Image Classification (MIC) is another widely studied area receiving attention from both medical and research communities. MIC is defined as categorizing images into different classes. MIC is either a binary class classification (e.g., malignant or benign) or multi-class classification (e.g., viral pneumonia, bacterial pneumonia, or Covid-19). Hospitals require experienced radiologists to classify medical images into various classes, which is time-consuming and vulnerable to instability and nonreproducible results. Effective application of MIC in hospitals could assist doctors in diagnosing diseases at a faster and more accurate rate [19]. Numerous studies involving the classification of medical images have been conducted for example, in [35], the authors performed 3-class classification to separate glioma, meningioma, and pituitary tumors using brain MRI images. In, [36], transfer learning techniques are employed to improve the accuracy of liver lesion classification using CT images. Yet another study [37], a novel CAD system to identify bacterial and viral pneumonia in chest radiography is proposed.

3. METHODOLOGY

Search Strategy and Data Collection

To attempt a reproducible and rigorous analysis, the data collection procedure is conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [38]. After defining the objectives of the meta-analysis, we selected the following databases for searching the relevant articles. PubMed, IEEE Xplore, Web of Knowledge, and ACM were searched for articles written in English. A major reason to select these databases is that the articles found in these web libraries are peer-reviewed. As DL techniques started to gain attention in 2012 [4], we targeted our search from 2012 to 2020. We prepared a customized query using the specific keywords employed for search such as “Transfer Reinforcement Learning”, “Federated Deep Reinforcement Learning”, “Federated Deep Transfer Learning”, “Federated Transfer”, “Deep Learning”, “Transfer Learning”, “medical img*”, “image analysis”, “radiography”, “x-ray”, “mammogra*”, “CT”, “MRI”, “PET”, “ultrasound”, “therapy”, “radiology”, “MR”, “SPECT”. This is done to capture a rich set of relevant papers from the above digital libraries and facilitate insightful text mining. After removing duplicate articles, we have a dataset of 8704 articles. Along with the current study, this article collection will be used in our future projects, especially in the Systematic Literature Review.

Data Analysis

From this rich text collection, we used our text mining process depicted in figure 1 to complete our analysis. We first pre-processed the abstract data to remove numbers, punctuations, and stop words. Next, we performed tokenization that can be used to prepare a word cloud and N-gram analysis. While a word cloud depicts the frequency of the words, N-gram analysis provides an overview of the specific keyword occurrences. Then we performed text-mining on the abstract data of the 8704 articles using VOS-viewer clustering. VOS-viewer tool uses a unified approach to mapping and clustering of bibliographic networks that is suitable to identify the main topics in the text data and then find relationships between the topics [39].

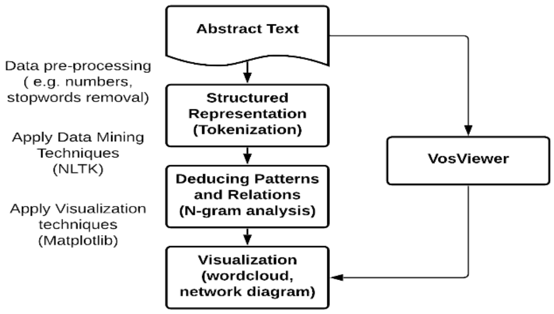

Figure 1. Our Text Mining process

4. RESULTS

Exploratory Data Analysis

A simple descriptive data analysis was conducted to gain insights into current trends using the gathered articles. This involved calculating 1. The number of articles published each year, 2. Word frequency of top ten words, 3. Word cloud, 4. Bi-gram analysis.

The number of articles published in each year from 2012 to 2020 is shown in figure 2. The articles published in the year 2015 have more than doubled. From the year 2016, the increase in publications is ascending followed by a slight drop in the year 2020.

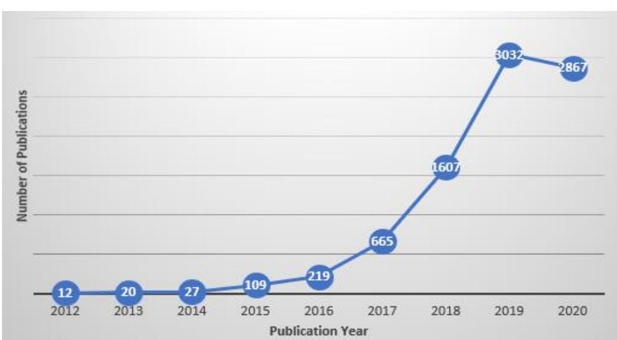

Figure 2. Number of articles published in each year

Figure 3 depicts the top 10 words in the corpus. Not surprisingly, ‘learning’, ‘deep’, and ‘neural’ are the topmost words with a frequency range between 7000 to 12000. Followed by ‘segmentation’, ‘convolutional’, and ‘classification’ with a frequency range between slightly lesser than 4000 to 6000. ‘Artificial’ and ‘intelligence’ also have a high frequency. The most frequent disease referenced in the corpus is ‘cancer’ with approximately 2000 counts. Examining single terms from the corpus does not provide rich insights on the themes, hence we further analyze the corpus using a word cloud.

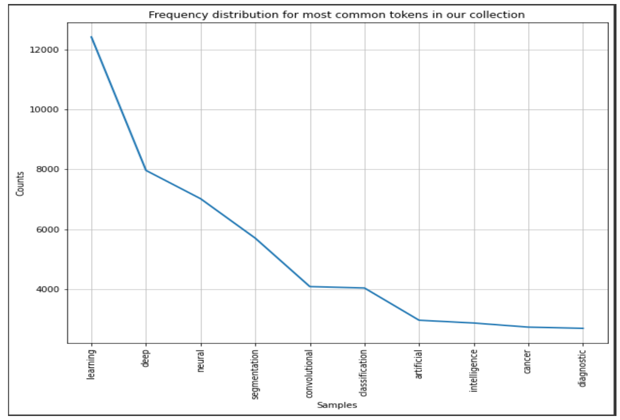

Figure 3. Top 10 most frequent words

Word clouds are a simple and fast approach that enables perceiving the most prominent terms from a large corpus of text data. They provide an overall sense of the data. The size of each term represents its frequency and/or importance. From the figure below, it is evident and quite expected that ‘deep learning’ is the most prominent word in our corpus. ‘Segmentation’ and ‘classification’ are the most dominant medical image processing tasks performed. Other image analysis tasks such as registration, recognition, and localization are less prominent. The most prevailing technique used for medical image analysis is the ‘convolutional neural network’. The most focused diseases are breast cancer, brain tumor, skin cancer, diabetic retinopathy, and Alzheimer's in decreasing order.

Although the word cloud provided an overall sense of the data, the specific learning modes – TL, RL, FL did not appear. To gain insights into the specific learning modes we performed a bi-gram analysis.

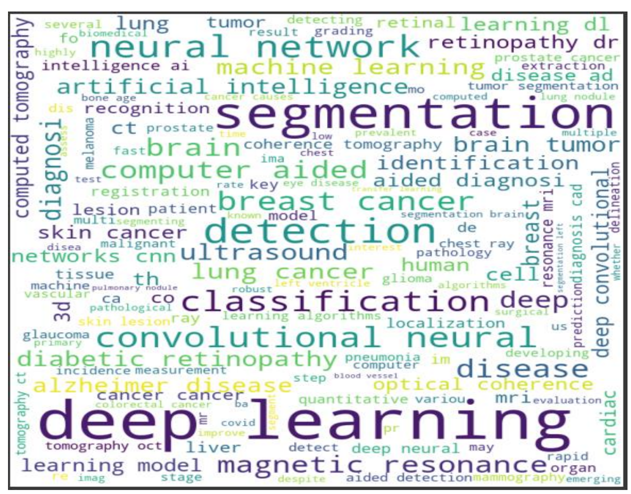

Figure 4. Word Cloud

Distribution of learning modes

The following figure represents the frequency distribution of learning modes from 2012 to 2020. As the figure depicts, Transfer learning is the most prevalent mode used since 2012. Reinforcement learning has also been used since 2012, however comparatively to a smaller extent. In 2014, reinforcement learning is almost as prevalent as transfer learning. While transfer learning usage increased in the following years, RL did not see much growth. Generative Adversarial Network is employed in medical imaging since 2018. A slight increase is visible in the following years. Compared to RL, GANs have been used more despite being proposed only in 2018. Federated learning is fairly new in the medical domain. It is barely visible in 2019 with a slight increase in 2020. While all the learning modes are individually researched in the medical image analysis field, the convergence of the learning modes is yet to ensue. Figure 5 depicts the distribution of learning modes that answers our RQ1 and RQ2.

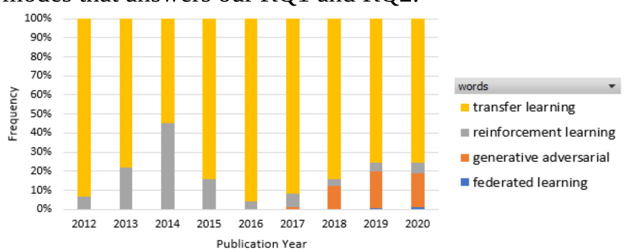

Figure 5. Distribution of learning modes

Research Themes

RQ3: What are the major themes in medical imaging literature? is answered with Figure 6, as indicated in the diagram below, six themes of medical image analysis studies are present. The Red cluster involved image segmentation studies, the green cluster involved studies with image classification, the dark blue cluster focused on evaluation procedures such as sensitivity and specificity, the light blue cluster involved Optical Coherence Tomography (OCT) studies, the yellow cluster involved MRI imaging studies, and purple cluster involved chest imaging studies. The bigger the size of the circle, the higher the occurrences of the word, the circles are connected with links, and the "total link strength" attribute in VOS-viewer is used to measure the relationship strength between the topics. Among the themes, the image segmentation has the highest link strength (total link strength = 9967) and is strongly connected to a variety of topics in all clusters especially Computed Tomography (CT) imaging, U-Net, and Magnetic Resonance Imaging. The research theme with the lowest link strength is OCT, it is (total link strength=1137) connected to both image segmentation and image classification themes.

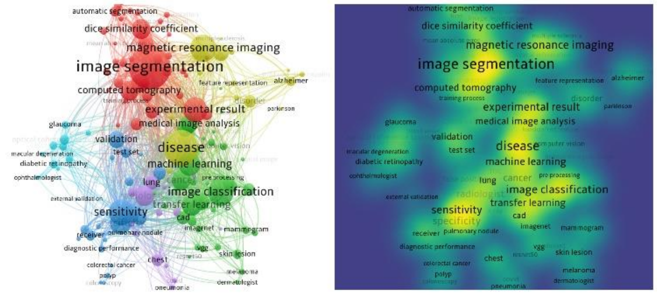

Figure 6. Clustering analysis using VosViewer (Left: Network Visualization, Right: Density Visualization)

Trending Topics

Figure 7 depicts the trending topics from the text data used from 2018 to 2020. The most recent words used in articles are colored in yellow and the oldest topics are colored in purple. The most recent medical image publications targeted towards U-net, GANs, covid, chest, f1 score, ResNet, DenseNet, VGG. Topics such as Support Vector Machine, Alzheimer's, mild cognitive impairment, heart, mammogram were published more in early 2018. Cancer, image segmentation, image classification, specificity, sensitivity, early detection are published more between 2018 and 2019.

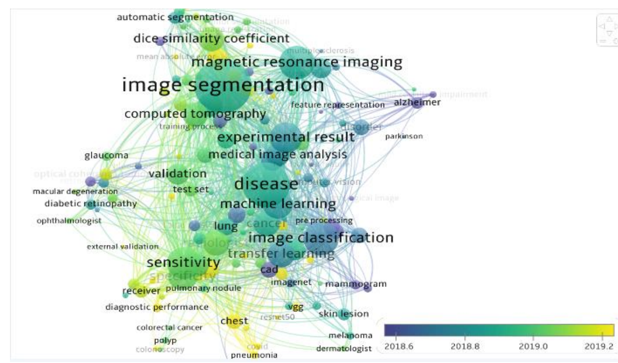

Figure 7. Trending Topics in Medical Imaging

5. DISCUSSION

In the current study, 8704 articles on medical image analysis from the four major databases are analyzed. The current literature on medical image analysis has six different themes - image segmentation, image classification, evaluation processes, MRI, chest imaging, and OCT imaging. Among the DL techniques, TL is the most prolific with a total link strength of 3984, the second most used learning mode is GANs with total link strength of 956. Unfortunately, RL and FL did not appear in the VOS-viewer clustering as their occurrences are significantly less when compared with the other two learning modes. However, from the n-gram analysis, it is evident that FL papers are being published from 2019 and publications increased in 2020 suggesting that it is fairly new and there is an opportunity to further investigate in this area. For RL, even though their overall occurrences are less than 100, the publication number increased consistently from the year 2012 to 2020 suggesting the trend will continue in the coming years.

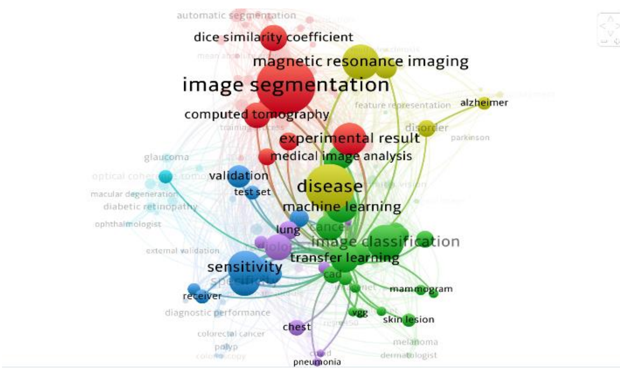

Figure 8. TL connections with other topics

Among the themes, the strong connection of image segmentation task towards U-Net, GAN, and MRI imaging suggests the preference of using U-Nets and GANs in segmentation processes for MRI imaging. On the other hand, the image classification task has a strong connection to topics such as TL, ResNet, DenseNet, ImageNet, AlexNet, VGG suggests the preference of using ImageNet pre-trained networks to apply TL in medical image classification tasks. In the evaluation procedures, sensitivity (total link strength = 6740) and specificity (total link strength = 5692) have strong links to all themes with topics such as dermatology, diabetic retinopathy, chest, radiologist, glaucoma, Alzheimer's, CT, OCT, MRI, etc. suggesting the importance of these two evaluation procedures in different areas of medical imaging.

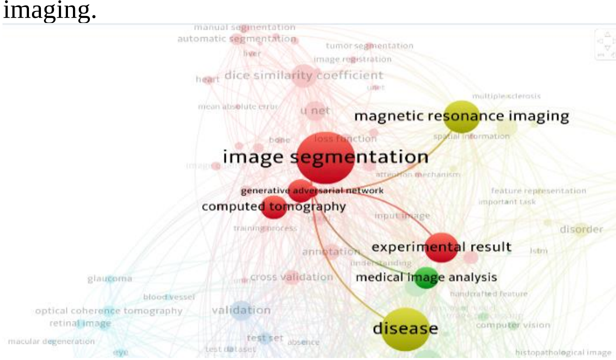

Figure 9. GAN connections with other topics

Among the learning modes, while GAN connections are limited with topics such as CT and MRI imaging modalities, TL is highly connected with various imaging technologies such as OCT, MRI, X-ray, mammogram, etc. suggesting the widespread applications of TL. Anatomical areas and diseases such as chest, lung, skin, cancer, Alzheimer's, pneumonia, covid, etc. are strongly connected with TL. This depicts TL is still relevant even with the most recent medical image analysis studies. GANs are depicted in a yellow circle in figure 7 suggesting it is a new learning mode with connections to MRI and CT. These results suggest that currently, TL is an established approach in medical image research while the other learning modes are fast-growing approaches with room to immensely expand individually and with hybrid approaches by combining with other learning modes.

In the network visualization diagram, the disease keyword is one of the most frequent keywords with a total link strength of 8922 and 1764 occurrences. The topics that are strongly connected to the disease include cancer, Alzheimer's, pneumonia, glaucoma, Parkinson's, covid, etc. Among these diseases, cancer has the highest total link strength of 3704 with 744 occurrences suggesting the wide applications of DL in cancer diagnosis. Covid, as depicted in figure 7 is the most recent application with a close relation to pneumonia. Also, several medical imaging areas such as diabetic retinopathy, radiology, CT, OCT occur more than 100 in the year 2020 depicting the fast-growing nature of DL applications in these areas.

6. CONCLUSION

To conclude, in this research study we conducted a literature analysis to explore the current DL trends in medical image analysis. We discovered that TL is the most highly researched learning paradigm followed by GANs. Nevertheless, RL and FL are also gaining attention from researchers in medical imaging. From a disease perspective, Cancer being a highly targeted disease and Covid diagnosis is the most recent DL application trend. Sensitivity and Specificity are the widely used evaluation metrics of DL models. Among the medical image tasks, image segmentation and classification are the major applications in different anatomical areas, classification tasks prefer using the CNNs AlexNet, VGG, DenseNet, GoogleNet, and segmentation tasks prefer using U-Nets. This finding is in alignment with previous comprehensive literature reviews in medical imaging analysis [8]. This study is not without limitations, we included only peer-reviewed online libraries in this study, databases such as ArXiv are excluded that contain pre-prints and early access articles that may not be considered peer-reviewed, but they contain the most recent technical developments. In the future, we plan to perform a systematic literature review to learn about the convergence of different learning modes and the issues affecting their convergence.

7. ACKNOWLEDGMENT

The authors would like to thank Chad Fenner, Kodey Crandall, and Andrew Behrens for peer reviewing the article.

8. REFERENCES

[1] S. Russell and P. Norvig, “Artificial intelligence: a modern approach,” Prentice Hall, 2002.

[2] A. Géron, *Neural Networks and Deep Learning*. O'Reilly, 2018.
[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
[5] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep Learning Applications in Medical Image Analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018, doi: 10.1109/ACCESS.2017.2788044.
[6] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *MEDICAL IMAGE ANALYSIS*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.
[7] H.-C. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 35, no. 5, SI, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.
[8] F. Altaf, S. M. S. Islam, N. Akhtar, and N. K. Janjua, "Going Deep in Medical Image Analysis: Concepts, Methods, Challenges and Future Directions," *arXiv:1902.05655 [cs]*, Feb. 2019, Accessed: May 07, 2020. [Online]. Available: http://arxiv.org/abs/1902.05655
[9] R. K. Samala, H.-P. Chan, L. M. Hadjiiski, M. A. Helvie, and C. D. Richter, "Generalization error analysis for deep convolutional neural network with transfer learning in breast cancer diagnosis.," *Phys Med Biol*, vol. 65, no. 10, p. 105002, May 2020, doi: 10.1088/1361-6560/ab82e8.
[10] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 3320–3328. Accessed: Aug. 29, 2019. [Online]. Available: http://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf
[11] F. C. Ghesu, B. Georgescu, S. Grbic, A. Maier, J. Hornegger, and D. Comaniciu, "Towards intelligent robust detection of anatomical structures in incomplete volumetric data.," *Med Image Anal*, vol. 48, pp. 203–213, Aug. 2018, doi: 10.1016/j.media.2018.06.007.
[12] I. Ali *et al.*, "Lung Nodule Detection via Deep Reinforcement Learning.," *Front Oncol*, vol. 8, p. 108, 2018, doi: 10.3389/fonc.2018.00108.
[13] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-Institutional Deep Learning Modeling Without Sharing Patient Data: A Feasibility Study on Brain Tumor Segmentation.," *Brainlesion*, vol. 11383, pp. 92–104, 2019, doi: 10.1007/978-3-030-11723-8_9.
[14] W. Y. B. Lim *et al.*, "Federated Learning in Mobile Edge Networks: A Comprehensive Survey," *arXiv:1909.11875 [cs, eess]*, Feb. 2020, Accessed: Sep. 13, 2020. [Online]. Available: http://arxiv.org/abs/1909.11875
[15] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," *arXiv:1602.05629 [cs]*, Feb. 2017, Accessed: Sep. 13, 2020. [Online]. Available: http://arxiv.org/abs/1602.05629
[16] H. H. Zhuo, W. Feng, Y. Lin, Q. Xu, and Q. Yang, "Federated Deep Reinforcement Learning," *arXiv:1901.08277 [cs]*, Feb. 2020, Accessed: Sep. 13, 2020. [Online]. Available: http://arxiv.org/abs/1901.08277
[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv:1512.03385 [cs]*, Dec. 2015, Accessed: Sep. 12, 2020. [Online]. Available: http://arxiv.org/abs/1512.03385
[18] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv:1409.1556 [cs]*, Apr. 2015, Accessed: Sep. 12, 2020. [Online]. Available: http://arxiv.org/abs/1409.1556
[19] P. Rajpurkar *et al.*, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," *arXiv:1711.05225 [cs, stat]*, Nov. 2017, Accessed: Apr. 10, 2019. [Online]. Available: http://arxiv.org/abs/1711.05225
[20] R. Godasu, D. Zeng, and K. Sutrave, "Transfer Learning in Medical Image Classification: Challenges and Opportunities," p. 7, 2020.
[21] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *NEUROCOMPUTING*, vol. 321, pp. 321–331, Dec. 2018, doi: 10.1016/j.neucom.2018.09.013.
[22] V. Joshua and S. Sivaprakasam, "Coronavirus: Bibliometric analysis of scientific publications from 1968 to 2020," *Med J Islam Repub Iran.*, p. 8, 2020.
[23] W. W. Hood and C. S. Wilson, "The Literature of Bibliometrics, Scientometrics, and Informetrics," *Scientometrics*, vol. 52, no. 2, p. 291, Oct. 2001, doi: 10.1023/A:1017919924342.
[24] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, Aug. 2010, doi: 10.1007/s11192-009-0146-3.
[25] Y. Yu *et al.*, "A bibliometric analysis using VOSviewer of publications on COVID-19," *Ann Transl Med*, vol. 8, no. 13, Jul. 2020, doi: 10.21037/atm-20-4235.

[26] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018, doi: 10.1007/s13244-018-0639-9.

[27] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.

[28] J. Gao, Q. Jiang, B. Zhou, and D. Chen, "Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview.," *Math Biosci Eng*, vol. 16, no. 6, pp. 6536–6561, Jul. 2019, doi: 10.3934/mbe.2019326.

[29] I. Goodfellow *et al.*, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2672–2680. Accessed: Apr. 10, 2019. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[30] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer Science & Business Media, 2010.

[31] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *arXiv:2001.05566 [cs]*, Nov. 2020, Accessed: Dec. 01, 2020. [Online]. Available: http://arxiv.org/abs/2001.05566

[32] Z. Jiang, H. Zhang, Y. Wang, and S.-B. Ko, "Retinal blood vessel segmentation using fully convolutional network with transfer learning," *COMPUTERIZED MEDICAL IMAGING AND GRAPHICS*, vol. 68, pp. 1–15, Sep. 2018, doi: 10.1016/j.compmedimag.2018.04.005.

[33] Y. Yang, C. Ye, X. Guo, C. Yang, and H. T. Ma, "Automatic Brain Mask Segmentation for Mono-Modal MRI," in *Proceedings of the 2020 10th International Conference on Bioscience, Biochemistry and Bioinformatics*, New York, NY, USA, 2020, pp. 124–128. doi: 10.1145/3386052.3386073.

[34] Z. Li, L. Zou, and R. Yang, "A Neural Network-based Method for Automatic Pericardium Segmentation," 2019. doi: 10.1145/3339363.3339372.

[35] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *COMPUTERS IN BIOLOGY AND MEDICINE*, vol. 111, Aug. 2019, doi: 10.1016/j.compbiomed.2019.103345.

[36] W. Wang *et al.*, "Classification of Focal Liver Lesions Using Deep Learning with Fine-Tuning," in *PROCEEDINGS OF 2018 INTERNATIONAL CONFERENCE ON DIGITAL MEDICINE AND IMAGE PROCESSING (DMIP 2018)*, 2018, pp. 56–60. doi: 10.1145/3299852.3299860.

[37] X. Gu, L. Pan, H. Liang, and R. Yang, "Classification of Bacterial and Viral Childhood Pneumonia Using Deep Learning in Chest Radiography," in *PROCEEDINGS OF 2018 THE 3RD INTERNATIONAL CONFERENCE ON MULTIMEDIA AND IMAGE PROCESSING (ICMIP 2018)*, 2018, pp. 88–93. doi: 10.1145/3195588.3195597.

[38] A. Liberati *et al.*, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration," *PLoS medicine*, vol. 6, no. 7, p. e1000100, 2009.

[39] L. Waltman, N. J. van Eck, and E. C. M. Noyons, "A unified approach to mapping and clustering of bibliometric networks," *arXiv:1006.1032 [physics]*, Jun. 2010, Available: http://arxiv.org/abs/1006.1032