Deep learning applications in analyzing ultrasound images of thyroid nodules: Protocol for systematic review

Yasaman Sharifi¹ , Saeed Eslami Hassan Abady², Morteza Danai Ashgzari³, Mahdi Sargolzaei^{4*}

¹PhD Student of Medical Informatics, Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

²Associate Professor, Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

³MSc Student of Computer Engineering, Department of Computer, Faculty of Engineering, Islamic Azad University, Mashhad Branch, Mashhad, Iran

⁴Assistant Professor, Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

Article Info

Article type:
Review

Article History:

Received: 2020-04-26
Accepted: 2020-05-17
Published: 2020-05-17

* Corresponding author:

Mahdi Sargolzaei

Assistant Professor, Department
of Medical Informatics, Faculty of
Medicine, Mashhad University of
Medical Sciences, Mashhad, Iran

Email:
sargolzaeim971@mums.ac.ir

Keywords:

Deep Learning
Medical Diagnosis
Ultrasonography
Thyroid
Systematic Review

ABSTRACT

Introduction: Ultrasound images are one of the main contributors for evaluating of thyroid nodules. However, reading ultrasound imaging is not easy and strongly depends to doctors' experiences. Therefore, a CAD system could assist doctors in evaluating thyroid ultrasound images to reduce the impact of subjective experience on the diagnostic results. With the best of our knowledge there is not any articles that actually provide a systematic review of deep learning application in analyzing ultrasound images of thyroid nodules and Hence, a comprehensive review of studies in this field can be useful, therefore the protocol of this systematic Review will be presented to reach this goal.

Material and Methods: This protocol includes five stages: research questions definition, search strategy design, study selection, quality assessment and data extraction. We developed search for relevant English language articles using the PubMed, Scopus and Science Direct. Inclusion and exclusion criteria were defined and flow diagram is conducted, from 623 studies retrieved, 27 studies were included, after quality assessment data was extracted based on defined categories.

Results: The result of this systematic review can help researchers with comprehensive view and the summary of evidence to present new ideas and further research and represent a state of the art in this field.

Conclusion: In this study a protocol was used for doing a systematic review on various deep learning applications in thyroid ultrasound such as feature selection, classification, localization, detection and segmentation. Articles were screened based on the following items: study and patient information, dataset, method, results and comparison method.

Cite this paper as:

Sharifi Y, Eslami Hassan Abady S, Danai Ashgzari M, Sargolzaei M. Deep Learning Applications in Analyzing Ultrasound Images of Thyroid Nodules: Protocol for Systematic Review. Front Health Inform. 2020; 9: 33. DOI: [10.30699/fhi.v9i1.220](https://doi.org/10.30699/fhi.v9i1.220)

INTRODUCTION

The thyroid gland or simply the thyroid is a little endocrine gland located in the front of the neck consisting of two lobes that may be affected by several diseases [1]. A thyroid nodule is a major clinical problem upon a world scale, which is reported as the first symptom of thyroid cancer. Thyroid nodule formally are defined as discrete lesions within the thyroid gland, radiologically distinct from surrounding thyroid parenchyma [2].

The prevalence of thyroid nodules in population is

increasing around the world especially in female patients. However, the estimated incidence of thyroid nodules is up to 67% of adults, but approximately 5–15% of these nodules are found to be cancerous [3]. Therefore, an accurate diagnosis of the malignancy of thyroid nodules is necessary to ensure the consequently appropriate clinical management [1], and reduce the significant medical health care costs of the fine needle biopsy (FNA) and/or surgery [4].

Ultrasound is one of the most common techniques and a key examination for the management, assessing and evaluating thyroid nodules [5]. There are many

advantages of ultrasound imaging such as safety, easily accessible, noninvasive and cost-effective. However, reading ultrasound imaging is not easy and strongly depends to doctors' experiences, levels, status and other factors [5].

Indeed, attaining a correct diagnosis of cancer in thyroid ultrasound image still remains a challenging task for radiologists [6]. Therefore, a Computer-Assisted Diagnosis (CAD) system could assist doctors in evaluating thyroid nodules ultrasound images to reduce the impact of subjective experience on the diagnostic results. These systems offer a second opinion for doctors by using image processing and machine learning techniques [5].

Deep learning is a growing trend of machine learning and an improvement of artificial neural networks (ANNs) through resembling the multilayered human cognition that making major advances in solving problems which are hardly solvable with traditional ANNs system [7, 8]. Compared to traditional machine learning, the deep learning approach allows automated features extraction from the input data [9, 10].

One of the main applications of deep learning is interpretation of medical images [11, 12], which specifically includes segmentation, diagnosis, classification, prediction, and detection of various anatomical regions of interest (ROI) [13].

There are a few studies that review the application of deep learning for medical diagnosis but not on ultrasound images [14, 15]. There are a number of studies which summarize the research of ultrasound CAD [14-17] but not on deep learning technologies. Huang et al. [18] presented an overview of the traditional ultrasound CAD systems and among 14 articles included, only 2 of them used deep learning techniques in thyroid ultrasound. Khachnaoui et al. [1] reviewed the research on ultrasound CAD systems based on deep learning on thyroid but only most recent research included and just 8 articles presented in detail.

This paper aims to address a protocol for systematic review on the applications of deep learning techniques in analyzing ultrasound images of thyroid nodules. To the best of our knowledge, this is the first study in the literature devoted to systematic review of deep learning applications in this field.

The rest of the paper is organized as follows: Section 2 presents the method of this protocol in five subsections including identify research question, search strategy design, study selection criteria, study quality assessment and data extraction. In fact, Section 2 forms the core of this paper since it details our proposed review protocol. Finally, in section 3 and 4 we draw result and our final conclusions.

MATERIAL AND METHODS

A review protocol is necessary to carry out systematic literature review. Indeed, we develop a systematic review protocol in order to facilitate the systematic review planning, and also ensuring the rigorousness and repeatability of our systematic review. The review protocol is based on the guidelines and structure suggested by Kitchenham [19]. As Fig 1 shows, the review protocol consists of the following five steps: research questions identification, search strategy design, study selection criteria, study quality assessment, and data extraction process.

In the first step, we form the research question to define the limit of the research and the questions that should be answered in the Review. The search strategy as the second step, narrows the search through construction of search string, as well as selection of different electronic databases as the source of the search. In fact, this step explores all published research papers from the selected databases that could be related to our review based on the keywords which are derived from existing studies. The third step involves filtering of the retrieved studies based on the pre-defined inclusion and exclusion criteria in order to get the most relevant papers. In the next step we consider quality assessment criteria (QAC) in order to assess the appropriateness of the selected studies. Finally, the last step involves collecting, summarizing and reporting the related information in order to answer the research questions.

Fig 1: Systematic review process [20]

Identify Research question

The aim of this systematic review is to address and assess the finding and results of deploying deep learning approaches for ultrasound images of thyroid nodules. In doing this, the following research questions are posed to guide the systematic literature review:

Which deep learning techniques have better performance in various application of analyzing ultrasound images of thyroid nodules?

What are their required pre-process procedures?

What are the sizes of their data sets?

Search strategy design

We have designed a comprehensive search string for relevant English language articles using the PubMed, Scopus, Science Direct to retrieve all presumably relevant studies up to August 2019 and Searches were re-run and updated in February 2020.

The search terms and constraints as our proposed strategy in PubMed is represented in Table 1. Note that using Mesh terms is needed to taking account other relevant and synonym key words. At first, a number of key terms related to the core concept of the research questions were posed and approved by the authors. Then, we build the query according to the PubMed search syntax and structure. Table 1 show our PubMed query box which is divided of 5 parts. First part (A), consists of the main keywords pertaining to Thyroid. In the second part (B), we searched for computer systems and algorithms related terms to identify studies that probably deployed deep learning techniques. In the third part (C), we searched for terms related to sonography. Within the fourth part (D), we applied keywords which are about the thyroid nodules/cancer. Finally, in the last part (F), we combine results of the above parts using the conjunction operator, I.e. AND.

Table 1: Search strategy in PubMed

Part	Field	Keywords
A	Title-abstract	"Thyroid" OR "Thyroid Gland"[<i>Mesh</i>]
B	All field	"Artificial Intelligence" OR "Deep learning" OR "Machine learning" OR "neural network" OR "CAD" OR "computer-aided diagnosis"
C	all field	"Ultrasonography"[<i>Mesh</i>] OR Ultrasound OR Ultrasonic OR sonography
D	all field	"Neoplasms"[<i>Mesh</i>] OR neoplas* OR cancer* OR tumor* OR malignan* OR benign OR "nodule" OR "nodules"
F	A AND B AND C AND D	

Study selection

Inclusion and exclusion criteria

To select only relevant papers for the subject of study, we defined the inclusion and exclusion criteria. Inclusion criteria of the study include the following:

Research or study has used CAD system based on machine learning technique;

The goal of the study should be in thyroid field;

In turn, excluded sources that met all the defined exclusion criteria had to:

Non original and Review study;

Non-ultrasound imaging technique have been used;

Do not use deep learning technique;

Do not focus on thyroid nodules/cancer.

According to the inclusion and exclusion criteria, two reviewers individually examined all titles and abstracts to separate related article to the purpose of the study. Discrepancies among the two reviewers were resolved by consensus involving a third reviewer.

After confirming the relevance, full texts reviewed by the same two reviewers. Discrepancies between these two reviewers were again resolved by consensus involving the third reviewer.

Flow Diagram of the research

This systematic review followed the PRISMA flow diagram and guideline [21], from identifying potentially related articles to the final included articles for the study. Fig 2 depicts the mentioned flow diagram, which consists of four main steps. At first, potentially related article based on the search strategy are identified from PubMed, Scopus, and Science Direct databases. In the next step, the screening process excludes some inadequate articles for our study based on their titles and then abstracts. In the third step, the eligibility of the articles was assessed after reading full text of the articles. Finally, the included articles were analyzed for the qualitative analysis and further classification regarding the aim of the systematic review.

Study quality assessment

We have adopted and modified a quality questionnaire that was proposed by Malhotra [20] to assess the credibility and strength of the included papers. Table 2 shows the quality assessment questions that comprise 13 questions, while each question has the following optional answers: "Yes" = 1, "partly" = 0.5, and "No" = 0. The final score of each study computed by summing up the scores of the answers in order to weight the studies'. Note that two independent researchers assessed the quality questions and consulted any discrepancy in their result with a third researcher to reach a consensus conclusion.

Data extraction

Data extraction was performed for all the 27 included studies by two reviewers independently and in duplicate using a predesigned table in Microsoft excel. For each study, a summary of the study and documented topics of interest as shown in Table 3 were extracted.

These two independent reviewers then implemented pilot-testing of the table on a random sample of five included studies until confirming a reliable data extraction. The calculated Kappa statistic [22] shows

the agreement of the reviewers on interpretation of the data and categories (kappa statistic=0.85).

Fig 2: Flow diagram of the review process

Table 2: Quality assessment questions

Q#	Quality questions	Yes	Partly	No
Q1	Are the aims of the research clearly stated?			
Q2	Are the variables related to dataset clearly defined? (#of patients, images, training and testing set)			
Q3	Is the data set size appropriate?			
Q4	Is the data collection procedure clearly defined?			
Q5	Are the DL techniques justified?			
Q6	Are the DL architectures clearly defined?			
Q7	Is the pre-processing procedure is clearly defined? (ROI detection and pre-process of images)			
Q8	Are the performance measures used to assess the result clearly defined?			
Q9	Are the results and findings clearly stated?			
Q10	Are the limitations of the study specified?			
Q11	Is the research methodology repeatable?			
Q12	Is there any comparative analysis conducted (Non-DL vs. proposed DL)?			
Q13	Is there any comparative analysis conducted (DL vs. proposed DL)?			

Table 3 shows the major elements of the extracted data, which were deemed to be critical to analyze for

this review. Note that the data table analyzed separately for each group of studies according to the deep learning application (feature selection, classification, localization, detection, segmentation). Moreover, several related finding to the study's purpose will be discussed and conclusions will be drawn.

RESULTS

In this review, all of the included studies have been published from 2017 to 2020, which indicates the growing of interest in the utilization of the DL techniques on thyroid nodules' ultrasound images in recent years. Based on the field of the DL applications, these 40 studies were divided into four categories: feature extraction (n=5, 13%), classification (n=16, 42%), detection (n=11, 29%) and segmentation (n=6, 16%).

None of the previous articles had a systematic review of all the deep learning applications in ultrasound images of thyroid nodules. To the best of our knowledge, this is the first study in the literature devoted to systematic review of deep learning applications in this field.

The results of this systematic review provide a comprehensive view of various application of deep learning in this field and we expect that our results will help researchers and also physicians and radiologists and other people who are interested in CAD tools based on deep learning by the summary of evidence to identify the state of the art and present new ideas and can help researcher to choose the right methods in future research.

DISCUSSION

In literature we can find many studies that address various applications of deep learning. Although, to the best of our knowledge none of them actually provide a systematic review on the quest that we study in this SLR.

In this study a protocol was used for doing a systematic review on various deep learning applications in analyzing ultrasound images of thyroid nodules such as feature selection, classification, localization, detection and segmentation and contributed an extensive literature review on the state of the art of the implementation of CAD systems based deep learning in this field.

CONCLUSION

With the best of our knowledge there is not any articles that provide a comprehensive review of deep learning application in analyzing thyroid ultrasound images. This article is protocol of review which helps researchers by summary of evidence to present new ideas and further research to reduce the health care

costs and patient's anxiety of the FNA or surgery.

Table 3: Extracted data from studies

Aspect	Variable	Description
General information	First author	
	Publication year	
	Title	
Patient information	Patient number	Patient number and image number
	Participant characteristics	Gender-age
Data Set	Type of dataset	Research dataset, online dataset
	Initial dataset	Number of images in initial dataset
	Augmentation Method	
	Final dataset	Number of images after augmentation
	Training, validation and test set	Number of various set
	Preprocessing methods	
Method and algorithm architecture	Application	Feature selection - Classification - Segmentation - Detection and Localization
	Approach	Deep learning- deep and traditional ML

Results and comparisons	Network architecture	
	Pre-trained	
	ROI detection algorithm	
	Comparison	Comparison to specialists/ Traditional techniques / Associated techniques
	Gold standard	Specialists/ Biopsy/Surgery
	Performance	Accuracy, sensitivity, specificity and etc.
	Main Result	

AUTHOR'S CONTRIBUTION

The authors agree on this final form of the manuscript, and attested that all authors contributed in the final draft of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this study.

FINANCIAL DISCLOSURE

No financial interests related to the material of this manuscript have been declared.

REFERENCES

- Khachnaoui H, Guetari R, Khelifa N. A review on deep learning in thyroid ultrasound computer-assisted diagnosis systems. International Conference on Image Processing, Applications and Systems. 2018: IEEE.
- Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, et al. Revised American thyroid association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2009; 19(11): 1167-214. PMID: 19860577 DOI: 10.1089/thy.2009.0110 [PubMed]
- Zhuang Y, Li C, Hua Z, Chen K, Lin JL. A novel TIRADS of US classification. *Biomed Eng Online*. 2018; 17: 82. PMID: 29914498 DOI: 10.1186/s12938-018-0507-3 [PubMed]
- Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *J Digit Imaging*. 2017; 30(4): 477-86. PMID: 28695342 DOI: 10.1007/s10278-017-9997-y [PubMed]
- Acharya UR, Swapna G, Sree SV, Molinari F, Gupta S, Bardales RH, et al. A review on ultrasound-based thyroid cancer tissue characterization and automated classification. *Technol Cancer Res Treat*. 2014; 13(4): 289-301. PMID: 24206204 DOI: 10.7785/tcrt.2012.500381 [PubMed]
- Li X, Wang S, Wei X, Zhu J, Yu R, Zhao M, et al. Fully convolutional networks for ultrasound image segmentation of thyroid nodules. *International Conference on High Performance Computing and Communications*. 2018: IEEE.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015; 521(7553): 436-44. PMID: 26017442 DOI: 10.1038/nature14539 [PubMed]
- Guan Q, Wang Y, Du J, Qin Y, Lu H, Xiang J, et al. Deep learning based classification of ultrasound images for thyroid nodules: A large scale of pilot study. *Ann Transl Med*. 2019; 7(7): 137. PMID: 31157258 DOI: 10.21037/atm.2019.04.34 [PubMed]
- Kumar A, Kim J, Lyndon D, Fulham M, Feng D. An ensemble of fine-tuned convolutional neural networks for medical image classification. *IEEE J Biomed Health Inform*. 2017; 21(1): 31-40. PMID: 28114041 DOI: 10.1109/JBHI.2016.2635663 [PubMed]
- Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: Review, opportunities and challenges. *Brief Bioinform*. 2018; 19(6): 1236-46. PMID: 28481991 DOI: 10.1093/bib/bbx044 [PubMed]
- Lee J-G, Jun S, Cho Y-W, Lee H, Kim GB, Seo JB, et al. Deep learning in medical imaging: General overview. *Korean J Radiol*. 2017; 18(4): 570-84. PMID: 28670152 DOI: 10.3348/kjr.2017.18.4.570 [PubMed]

12. Suzuki K. Overview of deep learning in medical imaging. *Radiol Phys Technol.* 2017; 10(3): 257-73. PMID: 28689314 DOI: 10.1007/s12194-017-0406-5 [[PubMed](#)]
13. Bakator M, Radosav D. Deep learning and medical diagnosis: A review of literature. *Multimodal Technologies Interact.* 2018; 2(3): 47.
14. Huang Q, Luo Y, Zhang Q. Breast ultrasound image segmentation: A survey. *Int J Comput Assist Radiol Surg.* 2017; 12(3): 493-507. PMID: 28070777 DOI: 10.1007/s11548-016-1513-1 [[PubMed](#)]
15. Jabarulla MY, Lee H-N. Computer aided diagnostic system for ultrasound liver images: A systematic review. *Optik.* 2017; 140: 1114-26.
16. Cheng H-D, Shan J, Ju W, Guo Y, Zhang L. Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognition.* 2010; 43(1): 299-317.
17. Sollini M, Cozzi L, Chiti A, Kirienko M. Texture analysis and machine learning to characterize suspected thyroid nodules and differentiated thyroid cancer: Where do we stand? *Eur J Radiol.* 2018; 99: 1-8. PMID: 29362138 DOI: 10.1016/j.ejrad.2017.12.004 [[PubMed](#)]
18. Huang Q, Zhang F, Li X. Machine learning in ultrasound computer-aided diagnostic systems: A survey. *Biomed Res Int.* 2018; 2018: 5137904. PMID: 29687000 DOI: 10.1155/2018/5137904 [[PubMed](#)]
19. Kitchenham B, Charters S. Guidelines for performing systematic literature reviews in software engineering. Keele University and Durham University Joint Report (EBSE 2007-001). 2007.
20. Malhotra R. A systematic review of machine learning techniques for software fault prediction. *Applied Soft Computing.* 2015; 27: 504-18.
21. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* 2009; 6(7): e1000097. PMID: 19621072 DOI: 10.1371/journal.pmed.1000097 [[PubMed](#)]
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33(1): 159-74. [[PubMed](#)]