VFAST Transactions on Software Engineering
http://vfast.org/journals/index.php/VTSE@ 2021, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 9, Number 4, October-December 2021 pp: 1-17

Deep Learning based Classification of Thyroid Cancer using Different Medical Imaging Modalities: A Systematic Review

Maheen Ilyas[1], Hassaan Malik[1,2,*], Muhammad Adnan[1], Umair Bashir[1], Wajahat Anwaar Bukhari[1], Muhammad Imran Ali Khan[1], and Adnan Ahmad[1]

[1]Department of Computer Science, National College of Business Administration & Economics Lahore, Multan, Pakistan
[2]Department of Computer Science, University of Management and Technology, Lahore, Pakistan
Corresponding Author: Hassaan Malik (f2019288004@umt.edu.pk)

ABSTRACT
Deep learning algorithms have achieved a tremendous triumph in task-specific feature classification. Deep learning methods are very much effective when a large amount of training data is scarce. It has been significantly applied for disease classification from medical imaging. The paper aims to identify and summarize the scenario of current research on thyroid cancer using deep learning methods through different medical imaging modalities which are found at present so that researchers become capable to select a useful and the most relevant approach which might be fruitful in dealing with thyroid cancer. This may also raise a need for more work out while dealing with future challenges. This Systematic literature review (SLR) has been presented by reviewing research articles published in well-reputed venues between 2017 to 2021. A comprehensive review was performed to appraise the deep learning approaches that have been applied in classifying a thyroid nodule disorder from different medical imaging modalities. The analysis is performed based on different parameters reported in selected research studies which include classification accuracy, true-positive (TP), false-positive (FP), true-negative (TN), false-negative (FN) sensitivity, specificity, and the area under the receiver operating characteristic curve (AUROC). A total of 2,149 research studies have been obtained by applying search queries in different journals' databases, out of them 40 papers have been selected for this SLR. Among them 22 studies have contributed sufficiently to the construction of the evaluation table which enabled the test process of methods of deep learning, having sensitivity varies between 75% to 100% (mean 89.50%) and specificity ranged from 64% to 100% (mean 84.4 %). The outputs revealed that the Convolutional Neural Network (CNN) has produced significant accuracy and has been extensively applied in the diagnosis of thyroid cancer by medical professionals. Furthermore, it is concluded that if the thyroid cancer exposure is inappropriate then it may restrict the deep learning mechanism and make its reliability challenge able.

KEYWORDS
Deep Learning, Thyroid Cancer, Nodules, Medical Imaging.

JOURNAL INFO
HISTORY: Received: October 11, 2021 Accepted: December 15, 2021 Published:.. December 31, 2021

1. INTRODUCTION

At present, thyroid cancer and the disorders of thyroid nodules are spreading rapidly. The treatment of this epidemic disorder is probably dependent on the improvement of medical technology [1]. During the past few years, significant cases of thyroid nodules have been reported, this is due to the continuous increase in the periodic dominance of ultrasound [2]. Thyroid nodules in the neck and chest region were detected by the tomographic scan, images developed by magnetic resonance, and Doppler. The ultrasounds having high resolution can detect nodules in about 50% of women under consideration [2].Thyroid nodules are more dominant in women and they have been increasing gradually in women with the age. A screening study was conducted through ultrasound in Germany on 96,278 respondents including both women & men having age between 18-65, In them, the percentage of thyroid nodules is 13% in women and 9% in men having age between 26-35, while the occurrence of thyroid nodules in the women and men having age 55 years or greater is 45% and 32% respectively, So from this, it is deduced that the thyroid nodules are most frequent in the people of older ages [3]. According to pathology, thyroid nodules are mostly benign, having a malignancy rate of about (5-7) %. A study showed that cancer which is an epidemic disease and having a high mortality rate account for (1-2) % of malignant nodules [4]. A method named Fine-needle aspiration (FNA) is used basically for predicting the malignancy risk occurring in the thyroid nodule, and on behalf of this estimation, the patients are being referred for surgery. Moreover, the Fine-needle aspiration is a broadly used and accepted technique for detecting the non-toxic nodules in the thyroid and hence leading the experts in deciding that whether the victim required surgery or not; and if surgery is required, (FNA) also helps in deciding the appropriate procedure for such surgery [5]. Besides this evidence, another study explained that FNA is the primary tool for the diagnosis and testing the thyroid nodules [6]. A study proved that most malignant nodules having a size of less than 1cm often cause less intense effects with positive test results [7]. Surgical resection is the technique used for the treatment of thyroid nodules, but a few of the post-operative problems encountered in this technique include hyperparathyroidism and laryngeal nerve paralysis, etc. Which may create some adverse effects on the patient's overall body [8]. Many researchers concluded that the patients that are suffering from thyroid nodule may need extra care, and this can raise a question on the overall effectiveness of ultrasound. For instance, try to find the answer to this query, he argued that ultrasound diagnoses for smaller malignant didn't have a significant approach and according to him this is the reason why such a patient needed extra care after diagnosing through ultrasound [9]. Based on the above reason, it should be necessary to look for some new technologies for making the ultrasound diagnosis of thyroid nodules, more standardized [10]. The rapid increase in the utilization of computer technology provided a big opportunity to achieve such milestones that can enhance the human approach in the different fields i.e. for the survival of life and in serving humanity, for instance, artificial intelligence (AI) and machine learning (ML) are examples of such technologies that can provide huge support to us in this regard. AI is the usage of machines that functioned like human beings such as learning and problem-solving. It deals with the intellectual abilities of machines by which one can achieve desired tasks with the help of robots in a smooth, reliable, efficient, and accurate way [11]. On the other hand, ML is another branch of AI that works on designed algorithms and different models of statistics which make the computer able to enhance the work performance of different tasks without writing the program openly [12]. For analyzing the images in medical diagnoses AI is one of the most significantly used technologies. AI can help to understand the digitalized images by providing much accurate information to doctors/physicians in making authentic and computer Aided Diagnoses (CAD). Ultrasound (US) is a commonly known imaging technique that is utilized in examining thyroid nodules [13], and authentic research proved the US imaging technology significant, in detecting the malignant thyroid nodules [14,15,16]. But the problem in using ultrasonography (US) is that it needs human aid i.e. the knowledge and experience of a radiologist is required while diagnosing through the US, so we can say that the ultrasound is operator dependent [17]. Researchers extracted hand-made classifiers for thyroid nodules with the aid of images taken from ultrasound, and the radiologists make specific decisions regarding the diagnoses, by utilizing their different expertise, knowledge, and experience about micro-calcification, shapes, margin, echogenicity, vascularity, and contents, etc. ML utilizes a series of algorithms and other features (e.g. features regarding texture, vascularity, and wavelet discretely transform features, etc.) from images taken through ultrasound [18]. Moreover, MKSVM and MALO classifiers if used for feature selection techniques from datasets, give more optimal conclusions compared to the existing ones. Under the light of examination, the study of [19] and [20] have experimentally proved the accuracy and appropriateness of these classifiers for thyroid diseases. Acharya et al. [18] also deduced through research that the accuracy of the classification that is made by ML is much better than the one given by radiologists through hand-made clinical classifiers. Moreover, ML also provides an additional non-clinical feature-based benefit i.e. it does not require any separate system with high specifications and its software can be installed in already existing systems serving in the clinics very easily. Based on such benefits and evidence, we can conclude that better results will be obtained if we automate the diagnosis of thyroid nodules through AI technology. Several studies have been executed in recent years, which aid in developing a CAD-based system that includes ML technology for categorizing the thyroid nodules in US imaging and also different data driving methods are introduced [18,21] and the results are much better as compared to those driven from US technology alone. The various tasks performed by the ML classifier involve the identification of thyroid nodules, classification of thyroid nodules, estimation of risk related to malignancy, and the identification of different types of malignancy [22].

The demand for diagnosis is rapidly increasing by the passage of time as the number of cases of thyroid cancer being reported is increasing gradually. Therefore, it is becoming a need of the hour that a more efficient and reliable method must be introduced that can produce more optimal and reliable results through analysis and data processing. By getting aid from the technology of artificial intelligence (AI) which is one of the emerging branches of computer science, one can improve the accuracy of diagnosis because artificial intelligence functions on a cognitive basis, and in this field, the computers are trained to impersonate the human minds, and surely this technique will help in providing an efficient diagnosis on one hand, and also give the stable diagnostic values on the other hand. In the recent studies, no evidence is present that can provide literature in which deep learning techniques are analyzed in the systematic study for diagnosing the thyroid nodule. Therefore, this research work aimed to put focus on this gap and also intended to summarize the latest, challenging, and pre-defined algorithms from the literature study to diagnose thyroid cancer and indicating the best-suited algorithms for disease diagnoses. For this purpose, a brief taxonomy is proposed by utilizing medical imaging, so the results of algorithms can be checked, and the accuracy of algorithms can be tested. Moreover, a model is also proposed for thyroid cancer disease diagnosis. We hope that this (SLR) work will prove to be helpful for human diagnosticians in classifying thyroid cancer disease by providing pre-defined learning practices related to a particular problem.

Here in this paper, we have categorized SLR as Section 1 contains Introduction, Section 2 consisted of methodology that was adopted for this SLR study, describing research queries, search strings, criteria of inclusion and exclusion of articles which were selected; Section 3 explains the brief results and the discussion that was made on the obtained results based on proposed model. The conclusion of this SLR is summarized in the Section 4.

2. MATERIALS AND METHODS

2.1 RESEARCH METHODOLOGY

The systematic literature review (SLR) provides help in identifying and analyzing the current research in the relevant domain of study. So, the purpose of this SLR is to identify the nearest accessible classifiers, finest methods, and datasets acceptability for the detection of thyroid nodules through deep learning techniques. By the classification of relevant studies, we provide scientific evidence for our findings. Kitchenham et al. [24] have recommended this systematic methodology for research and a similar method is used by [24] in prime studies.

2.2 REVIEW PROCEDURE

Establishing a proper review procedure is the key step for initializing any systematic literature review. The SLR procedures give support in identifying study strategies that can serve as a searching tool for extracting the data of interest from relevant literature. This procedure includes research questions, information sources, evaluation such as conference, journal and developing search strings, checking requirements for exclusion and inclusion. The full SLR process for this research is shown in Figure 1.

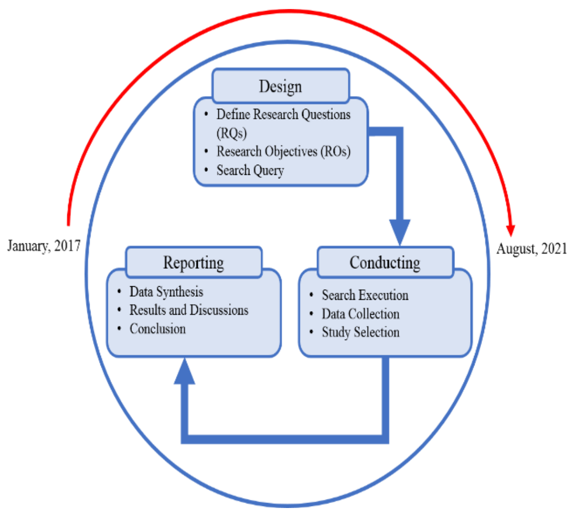

Figure 1. Systematic Literature Process

2.3 RESEARCH QUESTIONS

The primary purpose of this study is to address the problem of thyroid nodule diagnoses and to suggest a better way to diagnose this dreadful disease such as the problems identified by Gharib et al., [25]. A systematic literature review (SLR) addressing three questions is designed here to fetch a descriptive view of the problem under consideration. These three queries along with their most certain and probable motivations are shown below in Table 1. These queries provide us an obvious way to categorize the past and current researches that are conducted, regarding the problem of thyroid nodules and also give us a brief knowledge about the problem which may provide help in future research in this area to suggest the most reliable remedies to diagnose the thyroid nodules.

2.4 RESEARCH STRATEGY

In this section, we looked for the studies that are conducted for providing a primary model for the classification of any disease by utilizing the modalities of different medical and health imaging. Moreover, many researchers compare and differentiate between the authenticity of diagnoses carried out through algorithms and clinical experts. Here we focused on eight databases of different origins includes IEEE Xplore, Wiley Library, ACM digital library, Ovid-Medline, Scopus, Springer Link, Science direct and Taylor & Francis online & conference research publications which are published between

Table 1. Research questions

NO	Research Questions	Motivation
RQ1	What kind of datasets of different medical imaging is used in the classification of thyroid nodules using deep learning methods?	To determine the kinds of datasets that are utilized in searching for the effective way to diagnose the thyroid nodules with Deep Learning-based technology.
RQ2	What are the types of algorithms and different ways of classification employed, in suggesting the effective way to diagnose the thyroid nodules using Deep learning technology?	To check and determine the different algorithms which are implemented in AI-based technology for the diagnoses of thyroid nodules.
RQ3	Which parameters are being set for determining the truthfulness and authenticity of algorithms, implemented by using Deep learning technology, for the diagnoses of thyroid nodules?	The objective of this query is to find out the validity and precision that algorithm offers in terms of sensitivity also said to be as true positive or (TP), rate of errors and their types also said to be false positive or (FP),) specificity also said to be as true negative or (TN) and false negative or (FN.

2017 to 2021 which are evident for diagnoses through medical imaging for any sort of disorders. In looking for the more relevant studies manual searching was also done, citations and bibliography of articles selected for the study were also considered so that any relevant study if missed during the searching process could be taken and added in the selected articles. The query of well-reputed databases which are mentioned above are being run and the set of various keywords for running the queries are shown below in Figure 2. The primary and secondary search keywords are being selected as a basic identifier for the deep learning process and to capture the different publications which identify the accuracy of diagnoses by the clinical experts. Besides this, some more keywords are selected for detailed overviewing and full coverage of the data sets. The data collected from the selected studies are kept safe as a piece of basic information and knowledge, and hence can be filtered by the criteria and requirements of the study we needed. All the keywords and their different combinations which are designed for the search string of all (8) selected databases are shown in Table 2.

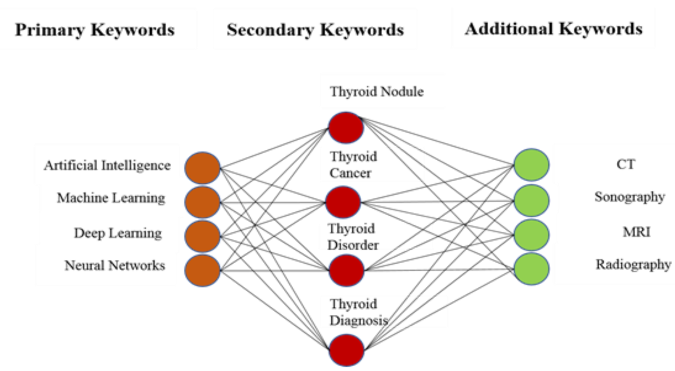

Figure 2. Search Strings Keywords.

2.5 STUDY SELECTION CRITERIA

The identification of the most relevant publication is an initial objective of any systematic study to commence the selection process. In our search criteria, if the same publication is obtained from one or more different databases it is accepted once only. All the selected publications were assessed and screened out by the author (s) personally by assessing the titles and abstracts of each publication. No limitations were put on the population being targeted. The criterion of the induction of the studies which were obtained was limited to search strings written in Table 2, a total of 2139 studies were taken from different databases, while 10 studies are taken from additional sources due to some validation threats.

All the papers selected for this systematic study are screened out based on the following exclusion criteria (EC):

EC1. Full-text articles which were irrelevant to the keywords are excluded as they were unable to meet the criteria.

EC2. Thyroid nodules that were diagnosed without the usage of imaging technology were not included.

EC3. The papers focusing on the segmentation instead of classification of the image were also excluded from the sample.

EC4. Publications that were based on samples other than human beings were also not included in our sample.

Figure 3 contained the results of the search process in which 40 articles were being selected out of 2149 filtered publications.

2.6 SCREENING PROCESS AND DATA ANALYSIS

The eligibility assessment was made by the two authors (MI and HM who independently perform screening on the abstracts and titles of the studies being searched and their non-consensus issue was being resolved by a third author (MA). We targeted only those studies which were on thyroid cancer. Contrary to this we have not set any limitation (s) on the target population, the outcomes of the interest, or the context for the model being used. Two authors (MI and HM) collected data separately using predefined data extraction sheets, the data was cross-checked, and disagreements are resolved through discussion, arguments, and support of a third, forth and fifth authors (MA, UB and WAB). And one author (MIAK) select and deselect the studies based on inclusion and exclusion criteria. We had contacted two other authors (MIAK and AA) to work further on the provided information. These two authors selected the papers by optimizing the objectives of the articles. We had not formally assessed the quality of the selected studies. Besides this, we had collected the binary diagnostic data where possible and build Evaluation Matrix. Our evaluation matrix consisted of true-positive or false-positive, true-negative, or false-negative conclusions, which were then utilized to calculate the sensitivity and specificity. We have done a Meta-Analysis of research studies for estimating the accuracy of deep learning algorithms and hence constructed an evaluation table. If selected studies provided different values for the same algorithms or vice versa, we had assumed them independent from one another. We have accepted such an assumption as we are interested to provide an overview of the results of selected research papers instead of performing a precise point estimate.

2.7 QUALITY ASSESSMENT

Quality assessment is a technique that is used to evaluate the quality of research. It is generally executed in a systematic study but sometimes it is carried out in mapping study as well. Here in this research quality assessment is carried out by some questions which are designed to review and evaluate the worth of all the 40 articles selected as a sample for this study. The quality assessment carried out by the author (s) of this paper has been shown in Table 3. The quality assessment is being carried out based on these three settled queries and the scores for each possible answer are also written against each query:

A) Are the algorithms of deep learning applied in this study for diagnosing thyroid disease? The scores assigned for two possible answers to this query are: “For Yes (+1)” and “For No (+0)”.

Table 2. Databases and Search Strings

Database	**Search Strings**
IEEE Xplore	((("Document Title": Deep learning OR transfer learning OR machine learning OR neural network) AND "Abstract": “Thyroid Nodules" OR "Thyroid disorders" OR "thyroid cancer" OR "thyroid diagnosis using AI") AND “Abstract": “CT” OR “Sonography”, “Radiographs” OR “MRI”) Publication Date: 01/01/2017 TO 08/18/2021
ACM Digital library	[[Publication Title: Deep learning] OR [Publication Title: transfer learning] OR [Publication Title: machine learning] OR [Publication Title: neural network]] AND [[Abstract:"thyroidnodules"] OR [Abstract:"thyroiddisorders"] OR [Abstract: "thyroid cancer"] OR [Abstract: "thyroid diagnosis using AI"]] AND [[Abstract: “CT” OR [Abstract: “CT”]OR[Abstract: “Radiographs”] OR [Abstract: “MRI”] OR [Abstract: “Radiographs”]]
Ovid-Medline	((Deep learning[Title] OR neural network[Title]) OR ((transfer learning[Title]) OR ((Machine learning[Title])AND("Thyroidnodules"[Title/Abstract]OR "thyroid disorders"[Title/Abstract] OR "thyroid cancer"[Title/Abstract] OR "thyroid diagnosis using AI"[Title/Abstract]) AND (“CT” [Title/Abstract] OR “Radiographs” [Title/Abstract] OR “MRI” [Title/Abstract] OR “sonography” [Title/Abstract])
Science Direct	Year: 2017-2021 Title, abstract, keywords: (Deep learning OR neural network OR transfer learning OR machine learning) AND ("Thyroid nodules" OR "thyroid disorders" OR "thyroid cancer" OR "thyroid diagnosis using AI") AND (“CT” OR “Radiographs” OR “MRI” OR “sonography”)

Wiley Library	"Deep learning OR neural network OR transfer learning OR Machine learning Title and "Thyroid nodules" OR "thyroid disorders" OR "thyroid cancer" OR "thyroid diagnoses using AI" OR "CT" OR "Radiographs" OR "MRI" OR "Sonography" in Abstract.
Springer Link	'(Deep learning OR neural network OR transfer learning OR machine learning) AND ("thyroid nodules" OR "thyroid disorders" OR "thyroid cancer" OR "thyroid diagnosis using AI" OR "CT" OR "Radiographs" OR "MRI" OR "sonography")' within 2017-2021
Scopus	TITLE-ABS-KEY(Deep learning OR neural network OR transfer learning OR machine learning) AND ("Thyroid nodules" OR "thyroid disorders" OR "thyroid cancer" OR "thyroid diagnosis using AI" OR "CT" OR "Radiographs" OR "MRI" OR "sonography)
Taylor & Francis Online	[All: deep]] AND [[All: learning] OR [All: neural]] AND [All: network] OR [All: machine]] AND [[All: learning] OR [All: transfer]] AND [All: learning] AND [All: network] AND [[All: "thyroid nodules"] OR [All: "thyroid disorders"] OR [All: "thyroid cancer"] OR [All: "thyroid diagnosis using AI"] OR [All: " CT"] OR [All: "Radiographs"] OR [All: "MRI"]] AND [Publication Date: (01/01/2017 TO 08/18/2021)]

B) Do the studies which are selected as samples provide any authentic solution in diagnosing the problem through medical imaging or not? The scores assigned for two possible answers to this query are: "For Yes (+1)" and "For No (+0)"

C) Are the 40 publications which are the samples of this study are published in some recognized journal or are they part of any conference of computer science?

The possible answers to this query for the publication that is published in some journals are being scored based on the quartile ranking (such as Q1, Q2, Q3, and Q4) of the journals in which the research is published, which is given by Journal Citation Reports (JCR), 2021. The scores for each quartile rank are given as under:

- If the paper is published in the journal of Quartile 1 then it is given a (+2) score.
- If the paper is published in the journal of Quartile 2 then it is given a (+1.5) score.
- If the paper is published in the journal of Quartile 3 then it is given a (+1) score.
- If the paper is published in the journal of Quartile 4 then it is given a (+0.5) score.

Moreover, the conference research papers are being ranked as CORE A, CORE B, and CORE C. The Scores for each conference rank are given as under:

- If the paper belongs to the CORE A rank, then it is given a (+2) score.
- If the paper belongs to the CORE B rank, then it is given a (+1.5) score.
- If the paper belongs to the CORE C rank, then it is given a (+1) score.
- If the paper belongs to CORE having no rank, then it is given a (+0) score.

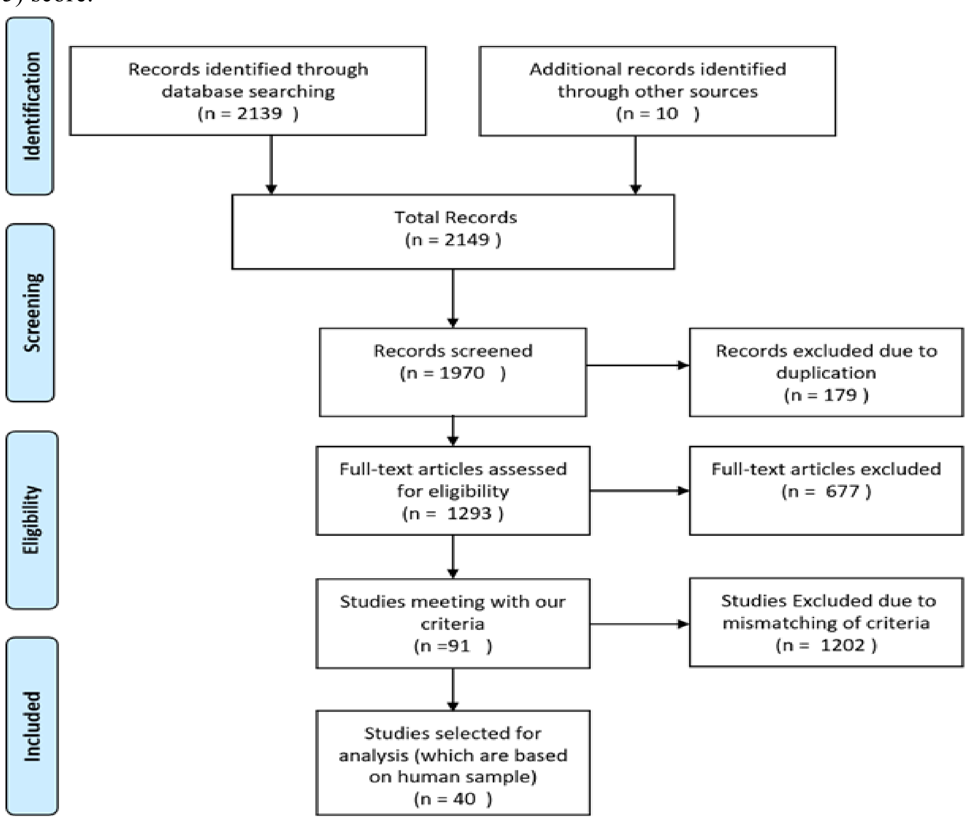

Figure 3. PRISMA diagram for Study Selection Criteria.

The point (c) mentioned in the above criteria is generated to enlighten the fact that the research papers that are published in journals are much more worthy than the papers which are presented in workshops, conferences, or seminars. So as the author (s) we believed that the publication of research work in a quartile-ranked journal is a very difficult and time taking phenomenon as compared to other means of publication. Moreover, after scores, we calculated descriptive statistics in table 3 that is based on the following given parameters:

(a) Number of research papers published in one year.

(b) Yearly averaged citations received by article, being reviewed.

3. RESULTS AND DISCUSSIONS

This section describes the answers to our three research questions that are written in Table 1. The scoring of Quality Assessment related to each preferred study is demonstrated in given Table 3. Out of the total selected articles, 29 articles (85%) scored above the average, 12% of articles obtained average scores and 3% of the articles got scores that were below the average. This Quality assessment can provide help to the health experts, readers, and researchers in the selection of relevant, useful, and appropriate studies, to diagnose the different diseases by medical imaging. Two different types of publication channels, which we have focused on in our study, are research conferences and journals. It has been concluded that 88% of the selected research papers belong to some journals and 12% of selected research papers were presented in some research conference.

Table 3. Quality Assessment

Evaluation Score						Citation Per Year						
Ref	**Publication Channel**	**a**	**b**	**c**	**Score**	**2017**	**2018**	**2019**	**2020**	**2021**	**Total Citation**	**Average Citation per year**
[26]	Journal	+1	+0	+1.5	2.5	0	1	1	7	2	11	2.2
[27]	Journal	+1	+1	+1.5	3.5	0	0	0	5	1	6	1.2
[28]	Journal	+1	+1	+1.5	2.5	0	1	4	5	2	12	2.4
[29]	Journal	+1	+1	+2	4	0	0	0	2	1	3	0.6
[30]	Journal	+1	+1	+2	4	0	0	3	8	1	12	2.4
[31]	Journal	+1	+0	+1.5	2.5	0	2	2	4	1	11	1.8
[32]	Journal	+1	+1	+2	4	0	0	0	3	1	4	0.8
[33]	Journal	+1	+1	+1.5	3.5	0	0	7	13	3	23	4.6
[34]	Journal	+1	+0	+2	3	0	0	5	18	1	24	4.8
[35]	Journal	+1	+0	+2	3	0	0	1	10	1	12	2.4
[36]	Journal	+1	+1	+2	4	0	0	8	19	1	28	5.6
[37]	Journal	+1	+1	+1	3	0	0	0	5	2	7	1.4
[38]	Journal	+1	+0	+2	3	0	0	0	2	2	4	0.8
[39]	journal	+1	+1	+2	4	0	1	2	8	2	13	2.6
[40]	journal	+1	+0	+2	3	0	1	1	7	1	10	2
[41]	Journal	+1	+1	+2	4	0	4	28	20	3	55	11
[42]	Journal	+1	+0	+0	1	0	0	0	1	1	2	0.4
[43]	journal	+1	+1	+2	4	0	0	2	12	1	15	3
[44]	journal	+1	+0	+2	3	0	0	1	14	2	17	3.4
[45]	journal	+1	+0	+2	3	0	0	0	0	1	1	0.2
[46]	journal	+1	+1	+0.5	2.5	0	2	3	7	1	13	2.6
[47]	Journal	+1	+0	+2	3	0	1	3	8	1	13	2.6
[48]	Journal	+1	+0	+2	3	0	5	15	20	1	41	8.2
[49]	journal	+1	+1	+2	4	0	0	14	11	1	26	5.2
[50]	journal	+1	+1	+2	4	0	0	13	13	2	28	5.6
[51]	journal	+1	+0	+2	3	0	1	7	24	1	33	6.6
[53]	journal	+1	+1	+1.5	3.5	0	0	3	1	2	6	1.2
[53]	Journal	+1	+1	+2	4	0	1	4	6	2	13	2.6
[54]	Journal	+1	+0	+2	3	0	2	9	25	6	42	8.4
[55]	journal	+1	+0	+2	+3	0	0	2	8	2	12	2.4
[56]	Journal	+1	+1	0.5	2.5	0	0	1	3	2	7	1.2
[57]	journal	+1	+1	1.5	3.5	0	0	1	2	2	5	1
[58]	journal	+1	+1	2	4	0	0	0	0	0	0	0
[59]	journal	+1	+1	2	4	0	0	0	0	0	0	0
[60]	journal	+1	+1	2	4	0	0	0	0	0	0	0
[61]	journal	+1	+1	2	4	0	0	0	1	0	1	0.2
[62]	Conference	+1	+1	+2	4	0	0	0	0	0	0	0
[63]	Conference	+1	+1	+2	4	0	0	0	0	0	0	0
[64]	Conference	+1	+0	+2	3	0	0	0	0	0	0	0
[66]	Conference	+1	+1	+2	4	0	0	0	0	0	0	0

3.1 SELECTION RESULTS

In this SLR we have identified 2149 records, from them 2130 were screened out (see Figure 3) and 40 studies contributed to this study, and the characteristics of whom are summarized in Tables 4, 5, and 6.

3.2 DATA EXTRACTION AND SYNTHESIS METHOD

The data extraction procedure was organized for providing adequate answers to the research questions mentioned in Table 1.

RQ1: What kinds of datasets of different medical imaging are used in the classification of thyroid nodules using deep learning methods?

Methods. Different types of images that are used in this study are JPEG, DICOM, Ultrasound images, and CT images. But most of the authors used US images for diagnosing thyroid cancer. The important and necessary step is to preprocess the data to fetch reasonable results. All the selected studies followed this step according to their requirement and in Table 4 the preprocessing technique of each selected study is mentioned. The significant and representative steps that are commonly used for preprocessing include segmentation, microcalcification, changes in composition, shape modification, Fixation of the margin, Transformation, etc. Analysis of data was carried out when the challenge was organized in 2018 JFR ("Journee's Francaises de Dataset is an essential component of deep learning which is being cited in journals and conferences. As an application of deep learning the Datasets of medical imaging have appropriate volume, adequate annotation, and proper re-usability. Every dataset of medical imaging is comprised of further related datasets, elements of data, and its attributes. Such a relation is termed an imaging examination. For medical imaging, Elements of metadata carried data formed by imaging modality. Data description depending upon the explanations and order that represent the components of the specific image. Table 4 consisted of details of datasets that described the diagnosis of thyroid cancer using deep learning Radiologie"). Several subjects were discussed in this challenge, but the authors decided to contribute to this challenge by working on the CT data of thyroid cartilages, one of them named Santin et al., [26] have used 515 CT data images in his work by transforming them into JPEG format. In the retrospective study of Zhao et al., [27], a new approach named DLRT ("deep learning Radiomics of thyroid") was utilized for the detection of thyroid nodules (benign and malignant) by using the DICOM image dataset taken from ACR ("American college of radiology"). From the clinical dataset, The US images of 40 patients were collected, who were under treatment of thyroid in some clinics during the last two years [28].

Table 4. Dataset Description

Ref	Datasets	Types	Instance	Format	Dataset Access
[26]	JFR dataset	JPEG	515	JPEG	Private
[27]	ACR dataset	DICOM	1629	DICOM	Private
[28]	Clinical dataset	Ultrasound	40	JPEG	Private
[29]	TIRADS dataset	Ultrasound	16946	PNG	Private
[30]	TIRADS dataset	Ultrasound	2836	JPEG	Private
[31]	Not Reported	Ultrasound	276	JPEG	Private
[32]	TCCC dataset	Ultrasound	370	PNG	Public
[33]	Hospital dataset	Ultrasound	62	JPEG	Public
[34]	STH dataset	Ultrasound	109	JPEG	Private
[35]	Clinical dataset	Ultrasound	106	JPEG	Private
[36]	PHC dataset	JPEG	589	JPEG	Private
[37]	PubMed dataset	Ultrasound	70	PNG	Private
[38]	ImageNet database	DICOM	1358	DICOM	Public
[39]	Tianjin Cancer	JPEG	11039	JPEG	Private
[40]	GPPH dataset	Ultrasound	2064	PNG	Public
[41]	Jinan Dataset	Ultrasonography	2064	JPEG	Private
[42]	Medical Center dataset	Ultrasound	1425	PNG	Private
[43]	SunYatsen dataset	JPEG	4738	JPEG	Private
[44]	Medical database	Ultrasound	1179	JPEG	Private
[46]	PubMed database	CT scans	2108	JPEG	Private
[47]	Peking Union Dataset	Ultrasound	4309	JPEG	Private
[48]	Clinical dataset	Ultrasound	117	JPEG	Public
[49]	BMU dataset	Sonographic	1077	PNG	Public
[50]	Hospital Dataset	Ultrasound	120	JPEG	Private
[51]	UPMC dataset	CT images	139	JPEG	Private
[53]	HOT dataset	3D scanners	20	JPEG	Private
[53]	Institutional dataset	JPEG	25	JPEG	Private
[55]	MIC dataset	Ultrasound	658	PNG	Private
[56]	PUMC dataset	Ultrasound	779	JPEG	Private

[57]	TIRADS dataset	Ultrasound	5278	JPEG	Public
[58]	Institutional dataset	JPEG	1817	JPEG	Private
[59]	Hospital data	Ultrasound	372	JPEG	Private
[60]	JFR dataset	JPEG	515	JPEG	Private
[61]	ACR dataset	DICOM	1629	DICOM	Private
[62]	Clinical dataset	Ultrasound	40	PNG	Private
[63]	TIRADS dataset	Ultrasound	16946	JPEG	Private

According to Kim et al., [35] 106 grayscale US images were collected from clinical patients by signing the written consent with them and few changes were applied in the by a single clinician for detecting the thyroid nodules. This study was accepted by the ethics committee of Jinan University, where a total of 4989 nodules were examined for identification of thyroid cancer using ultrasonography images [43]. Wang et al., [46] collected 4738 JPEG images from the cancer center of Sun Yatsen University for image recognition of thyroid carcinoma using deep learning methods. Ethical consent was acquired for this analysis, validation and training data for this analysis were gained through the medical database from Jan 2011 to 2016 [47]. Song et al., [49] used clinical image datasets taken from PUMCH ("Peking Union Medical College Hospital") after the approval of a senior doctor. Yoo et al., [50] a radiologist with having specialization with 10 years of experience in thyroid imaging evaluation, utilized clinical US images and set their ROIs manually as per the requirement.

RQ2. What are the types of algorithms and different ways of classification employed, in suggesting the effective way to diagnose the thyroid nodules using Deep learning technology? It was observed that out of 40 articles that were used in this study 29 articles were based on retrospective data while the remaining 11 studies consisted of prospective data. Moreover, dimension of images. Song et al., [39] used 3158 DICOM or Tiff images were taken from the ImageNet database and DICOM images were cropped into squares (299 x 299 pixels) 5 papers contain data that was originated from some open-access databases. There are two main processes, which are involved in deep learning when we consider it for diagnostic studies. The first process is based on classification which can be done by reducing the outcomes (diagnostic techniques) by matching the results with some specific or standardized outcomes. While the second process that is involved in deep learning belongs to the validity and authenticity of the data (medical data and images) obtained from various sources and is being utilized for diagnostic purposes. Besides this deep learning is also involved in dietary assessment. When considering the diagnosis of a particular disease it has been proved that the technique of deep learning is effective in many ways. Concerning the domain of deep learning a detailed review of our selected research studies is shown in Table 5. The research papers based on different kinds and several algorithms are also being presented in Table 5. Moreover, the algorithms employed of a different method in these selected papers and different deep learning architectures are presented in Figure 4.

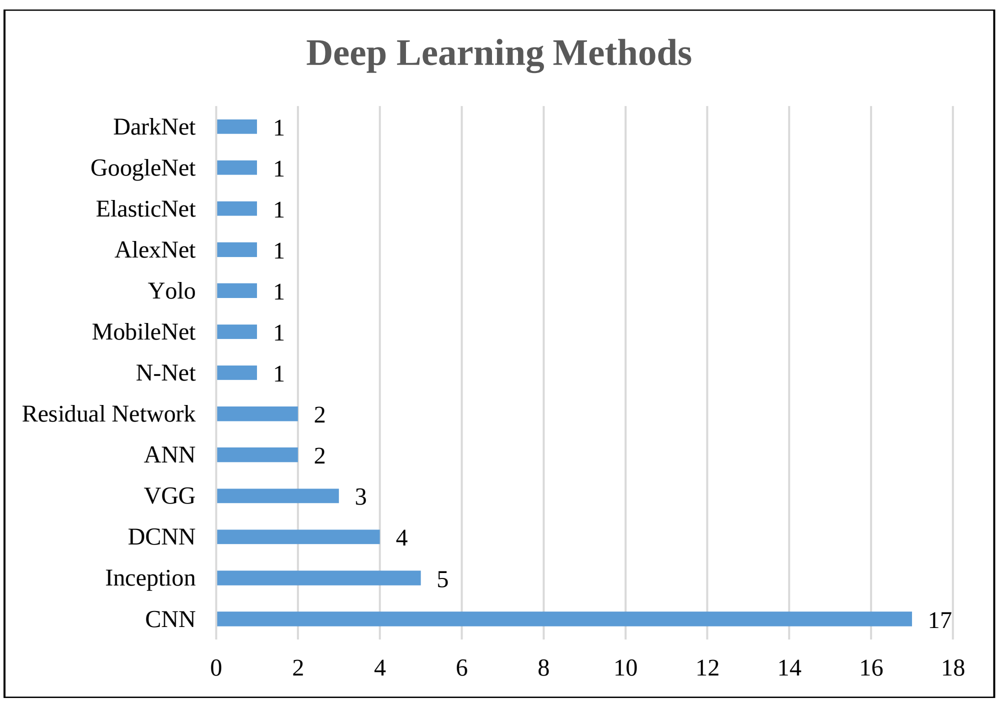

Figure 4. Synthesis of Reviewed articles by type of deep learning methods

RQ3. Which parameters are being set for determining the truthfulness and authenticity of algorithms, implemented by using Deep learning technology, for the diagnoses of a thyroid nodule?

The precision and accuracy of all the studies have been calculated through the aid of an evaluation matrix including sensitivity & specificity, precision & accuracy, and AUC (area under the curve)

Table 5. Different Deep Learning Architectures, Layers, and Data source of selected 40 studies

Ref	Architecture	Algorithms	Layers	Training Images	Gender (Male=M, Female = F)
[26]	VGG16	CNN	16,4,50	244	M= 300, F= 515
[27]	DLRT,TL	CNN	4	1097	1629 including M & F
[28]	SVM,RF,SFTA	SFTA	3	Not reported	M = 11, F = 29
[29]	Res-Net50	DCNN	50	4078	16946 including M &F
[30]	Inception-v3	CNN	48	1162	Not reported
[31]	Res-Net-50 v2	CNN	24	Not reported	M = 53, F = 223
[32]	RF,SVM	CNN	3	Not reported	370 including M &F
[33]	CNN	CNN	Not Reported	Not reported	M =12, F = 60
[34]	CAD	CAD	Not Reported	Not reported	M = 23, F = 53
[35]	Fisher exact test	McNemar	3	Not reported	M = 29, F = 77
[36]	Deep Net	CNN	16,19	439	M = 93, F = 426
[37]	ANN,SVM	ANN	3	Not reported	Not reported
[38]	Not reported	ML	Not Reported	Not reported	Not reported
[39]	Inception-v3	DLA	48	1138	1358 including M &F
[40]	SVM	CNN	12,3	Not reported	Not reported
[41]	Darknet-19	DCNN	50,19	17627	M = 577, F = 2124
[42]	RF	MLA	Not Reported	1240	2064 including M &F
[43]	LR, AdaBoost	CNN	3	1238	M = 695
[44]	CNN	TIRADS	Not Reported	1325	F = 1026, M = NR
[46]	RCNN	RCNN	Not Reported	196	M = 54, F = 253
[47]	Elastic Net	RF,KNN	Not Reported	700	1179 including M &F
[48]	CAD,GLCM	CNN	Not Reported	Not reported	2108 including M &F
[49]	GoogleNet	MC-CNN	16,8	6228	4309 including M &F
[50]	CAD	ANN	3	Not reported	M = 10, F = 40
[51]	BLR	CNN	5	Not reported	M = 795, F = 79
[53]	Not reported	CNN	Not Reported	Not reported	120 including M &F
[53]	LVM	SVM	12	Not reported	139 including M &F
[55]	Gaussian kernel	CNN	Not Reported	Not reported	M = 3, F = 46
[56]	VGG-f	CNN	8	1080	Not reported
[57]	ResNet/ImageNet	RCNN	Not reported	2461	Not reported
[58]	Inception-V3	CNN	3	679	Not reported
[59]	Inception-V3	CNN	Not reported	Not reported	Not Reported
[60]	CAD	CNN	Not reported	Not reported	M = 92, F = 288
[61]	CAD	CAD	Not reported	4916	M = 83, F = 289
[62]	CNN-F	CNN	3	Not reported	407 including M &F
[63]	Bi-LSTM	CNN	3	5436	Not reported

Moreover, the reliability of all classifiers was judged based on these attributes and parameters. In Table 6 the performance summary of a matrix has been presented. It was noticed that 22 research papers that are part of this study have given sufficient data that help in calculating the evaluation matrix. In deep learning algorithms [26, 27, 30, 31, 34, 35, 36, 37, 40, 41, 42, 44, 47, 50, 53, 62], the sensitivity ranged between 75.0 % to 99.9 % (mean 89.50 %) and specificity ranged between 64.0 % to 100.0 % (mean 84.4 %). Santin et al., [26] have used the JFR dataset and applied the CNN algorithm to it and reported 83% sensitivity and 64% specificity in the conclusion of his study. Prochazka et al., [28] used a clinical dataset and proved the accuracy rate of the SVM as 94.64% in his study. Jeong et al., [34] have used dataset and applied ResNet-50 on this dataset and presented 90.3% accuracy in his analysis. Data were obtained from a single tertiary hospital and apply this data on CAD to obtain accuracy rate. Yeon Ko et al., [36] have used a healthcare dataset and reported the sensitivity and specificity (80.0% and 90.0%) respectively. Data was gathered from the local hospital and used for applying different method like SVM and reported sensitivity and specificity was 94.4% and 89.3% respectively [40]. Tobriner et al., [44] used the AMC dataset for applying methods to find out the AUC (Area under the curve) that was 0.91% in their analysis. Data was gathered from some clinics and applied to some algorithms which showed a compromising accuracy rate of 88.6%.Based on the above literature, we have proposed a model in figure 6, and taxonomy after observing different experimental works. Our proposed model i.e. figure 6 can be applied to all the categories of the thyroid disorders which are presented in taxonomy (see figure 7).

Table 6. Evaluation matrix for thyroid cancer classification

Source	Classifier	Sensitivity	Specificity	Accuracy	AUC
[26]	CNN	83%	64%	75.8%	0.72%
[27]	CNN	90%	82%	Not Reported	0.96%
[28]	SVM & RF	89.9%	92.46%	94.64%	95.2%
[29]	VGG16, ResNet50	Not Reported	85.2%	0.83%	89.2%
[30]	Inception v3	93.3%	87.4%	90%	92.1%
[31]	Resnet50	90.46%	89.91	90.3	90.0%
[32]	KNN, SVM, RF	95%	97.8%	98.1%	0.850%
[33]	CNN	21.4%	81.3	67.7	78.2%
[34]	CAD	75.0%	80.4%	76.0%	0.86%
[35]	CAD	81.4	68.2%	73.4%	Not Reported
[36]	CNN & VGG16	84.0%	90.0%	86.0%	0.850%
[37]	Google Net	99.10%	93.90	98.29%	Not Reported
[39]	Inception-v3	94.0%	90.0%	86.0%	0.850%
[40]	CAD, SVM	94.4%	89.2%	98.2%	0.90%
[41]	DCNN, ResNet-50	93.4%	86.1%	92.6%	0.947%
[42]	RF	88.1%	84.2%	84.6%	0.94%
[44]	CNN	93.3%	64.7%	76.4%	0.91%
[47]	Elastic Net	90.07%	94.9%	96.7%	0.954%
[49]	MC-CNN/VGG-16	0.98%	0.98%	93.8%	0.98%
[50]	CAD	92.0%	85.1	88.6%	74.6%
[51]	BLR	41.7	93.7	79.8%	0.777%
[53]	CNN	97.6%	98.7%	98.3%	96.9%
[56]	CNN/SVM	96.4%	83.1%	92.5%	0.881
[57]	RCNN	92.5%	94.7%	96.3%	93.4%
[58]	CNN	99.1%	92.1%	96.7%	94.6%
[60]	CNN	99%	97.3%	Not reported	0.646
[61]	CAD	97.6%	21.6%	49.6%	Not Reported
[62]	CNN-MPR	86.99%	92.35%	90.17%	Not Reported
[63]	Bi-LSTM	0.939%	Not mentioned	0.8618%	0.9361%
[66]	CNN/VGG	0.722%	0.970%	0.804%	96.4%

The proposed taxonomy of the earlier work on thyroid classification is shown in Figure 7. Taxonomy is developed to show the clean picture of thyroid nodules classification from recent studies. There was a total of ten attributes based on which taxonomy is developed. The attributes are selected as they were brewing in the studies being surveyed. These ten attributes include adeno carcinoma, papillary carcinoma, follicular carcinoma, oxyphilic carcinoma, medullary carcinoma, adenoma, follicular, hurthle cell, medullary, and thyroid parenchyma.

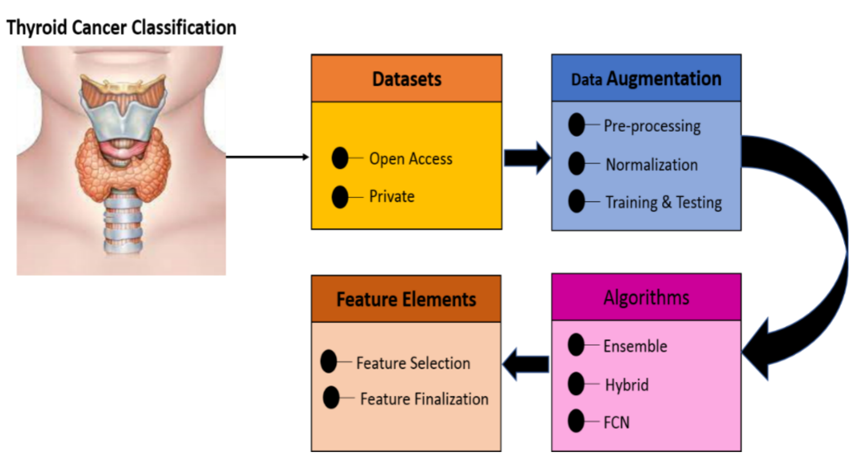

Figure 6. Proposed Model for the classification of thyroid cancer.

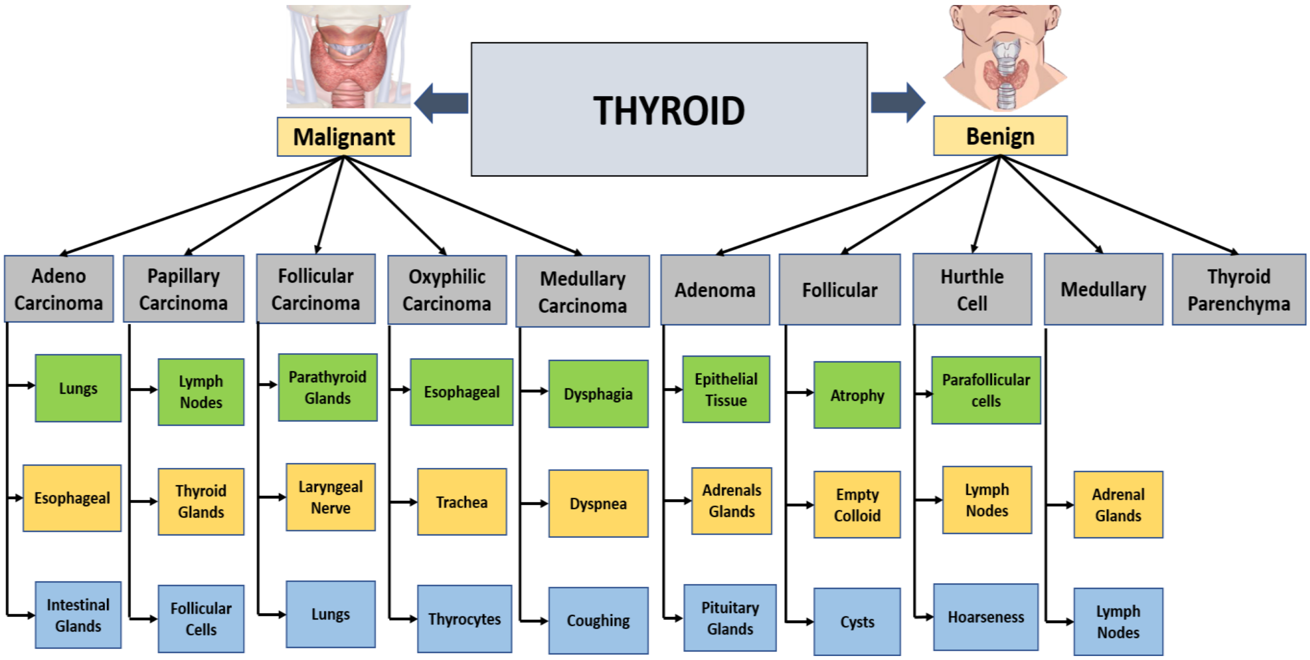

Figure 7: Taxonomy of Thyroid Disease

3.3 DISCUSSION

In medical imaging. this is the first study that is based on a systematic approach for the diagnoses of thyroid nodules. To conduct this research the selection of all studies has been made with great care by analyzing the performance, reporting, and validation of algorithms used in them. Most of the studies included in the sample lack preprocessing like [28], [32], [33], [39], [42], [44], [48], [49]. Contrary to this, few studies remained inadequate as they considered only a few types of thyroid nodules like [36], [37]. Moreover, some studies do not show generalized results due to the sampling bias they faced in collecting samples [39], [48], [50]. To comprehend the results and findings of this SLR study, a brief taxonomy is proposed as seen in Figure 7. The Taxonomy consists of the categorization of thyroid nodules in further subcategories. And then the final model is proposed for the diagnoses of thyroid diseases. The above model proposed in Figure 6 will prove to be an effective model for the diagnosis of diseases in medical imaging. The model consists of five main attributes including datasets collection procedures, distribution of data sets into a training dataset and testing dataset, utilization of augmentation technique for datasets, employing deep learning algorithms for fine-tuning of datasets, and measuring the accuracy of diagnoses. Two kinds of datasets are utilized in this study one is open access which indicates the dataset having publically accessible as a benchmark dataset for medical imaging and the other one is the non-publically available dataset. Different researchers addressed different techniques for feature selection of clinical datasets such as Fisher's, Discriminant ratio, Kruskal Wallis analysis of variance, and ROI. But few studies in which feature selection is done by using ROI are mentioned in our references as [27], [29], [36], [37] and other studies used Fisher's Exact for feature selection like in [30], [34], [35], [50], [53] in our references. The data sets having low contrast are being managed properly by applying the data augmentation technique. The various kinds of transforming procedures are also applied on datasets such as horizontal and vertical flips, zooming, and shifting so that a better accuracy level would be achieved. Moreover, the training and testing images of datasets are also defined at this stage. After all of this, the appropriate and most suitable pattern is being filtered out and selected. Furthermore, different types of algorithms are being applied to analyze the diagnostic accuracy of the deep learning technology. Various methods of deep learning technology used in this study, are based on transfer learning mechanisms whereas few techniques are based on ensemble methods, while some methods utilized here consisted of hybrid models or convolutional neural network models [66]. The result obtained through these techniques has been evaluated effectively and efficiently so that the best technique is identified and selected as shown in Figure 6.

The taxonomy designed here provides help to categorize the nodules into benign and malignant, to classify the disease in binary form, this research utilized deep learning techniques. When the thyroid disorder is being diagnosed as malignant, it is studied further to indicate its proper type. Malignant has four basic forms which are Adenocarcinoma, Superficial-spreading, medullary-carcinoma, follicular-carcinoma, and papillary-carcinoma. Deep learning techniques are significant in detecting a particular kind of carcinoma. All the types of malignant have variant structure, allocations, and shape, sizes, and color combinations. These four forms of malignant are being categorized further and they have quite horrible effects on the modality of humans. In figure 7 categories of thyroid diseases are mentioned, and all of these categories are taken from 40 selected studies, and few prominent out of them are mentioned here by their reference numbers according to our paper as [30][34][35][37][43][50], etc. The datasets which are used in such selected studies are labeled and verified by some clinical experts which are then trained on deep learning algorithms obeying all the rules and regulations. Contrary to this whenever the disease is detected as benign, it is then categorized into five various kinds which are follicular and hurthle, adenoma, thyroid parenchyma, and medullary. All of the types mentioned here are considered non-cancerous types of this disorder. They appear on the skin and are quite similar to malignant. Figure 7 contained the summary of results which are extracted by the established taxonomy which is designed to identify the thyroid nodules utilizing the deep learning technique.

3.4 PRINCIPLE FINDING

In this systematic study, we have selected 40 articles from the field of deep learning for diagnosing the disease thyroid cancer. By carefully selecting the papers of interest through reporting them transparently we have tried to find out the performance and validity of algorithms in diagnosing a thyroid disease (see Table 6). Results of this systematic study revealed that the performance of deep learning algorithms is admirable in diagnosing the disease of interest. The CNNs approach is used and it has proved the factual isolated behavior based on the provision of datasets under examination. On the other hand, the neural network approach needs a bulk of data sets for training procedures, and it is feasible for applying on models of pre-trained subjects in rare cases [67]. By combining the emerging technologies with pre-trained models e.g. if we talk about the cloud-based platform for data distribution, it would help enhance the performance more than single clinic operations [68]. The selected articles show the effectiveness of the technique of deep learning in disease diagnoses through medical imaging. It was seen that the computerized methods give more vigorous and efficient results after identifying the symptoms of a particular disorder like one (benign and malignant) which is discussed in our study. Through the usage and implementation of deep learning algorithms for diagnosing the diseases, we can reduce the burden from the clinical professionals, and hence the efficiency and effectiveness of diagnoses can be enhanced. There are more prominent modalities for image detection in this prospect. To diagnose thyroid cancer clinically various guidelines are being revealed which may be helpful in the selection of thyroid nodules that are needed to be diagnosed first. The article which is reviewed during this study proved that the success of any model depends on diagnostic results. Indeed, the understanding of the diagnostic outcome of the disease depends on the validity and authenticity of the suggestions. Recursive phenomena were applied as criteria for diagnoses in real and true situations and this was approved by experts having medical backgrounds. Besides this, the learning abilities of clinical experts often lead towards the identification of other prospects and also leads towards feasibility check for diagnostic phenomenon and this thing was added to their clinical experience whenever they become full professionals. Due to this reason, they like to adopt such models in their pre-training phase or learning phase which are being controlled and evaluated by experienced health professionals. Therefore, we cannot ignore the connection and relation between algorithms and medical experts. Thyroid cancer detection in early stages leads towards successful cure therefore it is quite essential to diagnose it as soon as possible. To get cytological consequences nowadays FNA is of wide acceptance as it is rapid, easy, and appreciable. However, diagnosis of FNA through biopsy gives 5% to 20% undefined and doubtful consequences and it needs more deep pathological examining [70]. In contrast to FNA classification of the image is less doubtful and less invasive and does not need any further deep examination of the health experts. Deep learning helps to improve the diagnoses of the thyroid by providing better accuracy of ultrasound images, it is also highly efficient in neural networks. Deep learning favors in better exploring of ultrasound images, it offers multi-dimensional features like the flow of blood and hardness of thyroid nodules. The system based on such intellectual diagnoses having multimode features for image detection has proved to be the developmental and innovative trend for ultrasound diagnoses. The technology of artificial intelligence revolutionized the field of diagnoses and its rapid progress gives an efficient exposure to diagnostic approaches. The deep learning techniques are fully dependent on human interactions in diagnosing disease, as results of diagnoses have to be implicated in real-world, therefore, they required the approval of some medical professionals. Deep learning approaches have contributed a lot in the diagnostic field by saving time and reducing fatigue both physical as well as cognitive in human beings [70] and resultantly it has improved the medical diagnoses of the experts and made their reports more reliable. Moreover, through the use of deep learning approaches less time is consumed in the diagnosis of the diseases which may result in less fatigue and reduced transitions of tension to medical experts from the clinicians.

Through deep learning the results of diagnoses are becoming more reliable and hence an enhanced and early health facility is granted to the patients after diagnoses. The interpretation of data is still becoming an important matter in AI. The emerging trends of reporting that addressed specific challenges of deep learning may provide new paths and routes towards future workings, which may contribute to providing better results, and more reliable outcomes are to be expected to reveal.

3.5 FUTURE SCOPE

With the collaboration of deep learning and clinical experts, more useful and challenge able outputs can be achieved in the future. Deep learning technology helps in achieving the most accurate results through continuous training and learning [70]. This may result from the effective and more accurate evaluation of medical images for disease diagnoses, from a future perspective. Our review of 40 articles concluded that the deep learning mechanism with its distinguishing features will contribute a lot to the improvement of the outputs of the medical professions, which human experts alone cannot acquire otherwise. Hence with these emerging technologies, the detection of thyroid nodules become easier in the future.

3.6 LIMITATIONS

Besides the benefits which are discussed in the previous section, AI is still a long-term procedure, so patience is required while dealing with it. For validating and verifying the system accuracy a bulk of data is required firstly. The data collection takes a long time, specifically when there will be the absence of appropriate standards for diagnostic. Furthermore, time and effort are required by the physicians so that they may become aware of the usage of different formulas and algorithms applied at the time of the modeling process. In the current scenario, a specific algorithm model is applied for a specific database. For improving the diagnostic concurrency of the database, it is essential to reorganize or adjust its algorithms or parameters. Contrary to this there may exist some undesirable flaws and problems in clinical applications. For example, there may exist some nodules which have disordered and irregular structures due to which unclear boundaries are evidenced in some areas but not in all areas, and due to this reason, the chances of misdiagnose become greater. To avoid this, misdiagnosis a descriptive scanning is required moreover an image having 3- dimensional structure may prove to be another solution to avoid the misdiagnoses.

3.6.1 LIMITATIONS THREATS TO VALIDITY

The approach followed for this systematic literature review was the one suggested by kitchenham [71] so that any partiality in selection could be avoided. All researchers belong to different backgrounds and communities so we have used various terms for a single topic, as per our knowledge we have used many different terminologies whom we are aware of and which are related to the diagnosis of the disease under consideration, so conclusions and findings of all published articles were inspected thoroughly and accordingly. As we have concluded our survey's findings until August 2021, so publications that appeared after that point of time are not part of this study. Therefore, by reflecting on this methodology, our study faced the following limitation.

3.6.2 SAMPLING BIAS

Initially, our research was based on published metadata including abstract, title, and keywords. Moreover, we retrieved data from the published papers that are written by good authors and also published through popular journals and conferences but there were chances that it might be overlooked twice as most of the authors often used the same data for different purposes.

3.6.3 REPORTING BIAS AND CONSTRUCT VALIDITY

To control the reporting bias and to check the construct to avoid misrepresentation of reporting, we have thoroughly understood the sample articles and repeatedly discussed every little matter among the researchers, so that we may answer every research question in a quite clear and understandable way. In short, our whole study was conducted in light of our research objectives and research queries so that all the findings and conclusions will be bridged efficiently.

3.6.4 DATA EXTRACTION AND INTERNAL VALIDITY

For ensuring the correct and coherent data extraction we have designed our methodology under the guidance and group discussion of all the experienced supervisors and authors of this article so that every possible human error or flaw regarding data extraction can be controlled and avoided. Our methodology was shown clearly in section II.

3.6.5 REPRODUCIBILITY

This study was replicated conveniently by using efficient selection or searching techniques. We have used automated means for the collection of data. But we have found it quite difficult and time-consuming mean for data collection as most of the libraries do not permit or forbade the public to retrieve their data through web restrictions, but few which permit or make their data exposed publically give limited access to their web. Some of the online libraries only allow their official partners or paying users to retrieve data from their online libraries I.e. IEE & Wiley, ACM give quite limited access in this regard. Therefore, resultantly we can say that automated means of data collection is no doubt quite a reliable method, but it required time and cost to be accomplished.

3.6.6 EXTERNAL VALIDITY

In our research, we only focus on published works and academic research channels, and it is a generalized survey that is concerned by peer review academic research works only. Every bit of information used or retrieved during this survey was properly discussed so that systematic errors are to be avoided and hence the quality of the survey should be maintained in a good manner.

3.7 CHALLENGES AND OPPORTUNITIES

For detecting the thyroid nodule through deep learning one can face many challenges and opportunities. In this section, we discussed those challenges in detail which are identified through the literature review.

3.7.1 DATASET VARIATION

For classifying the thyroid nodules many different datasets are analyzed in detail. Few datasets are easily available for the whole public, but some are not. The number of images is different in each dataset. Besides this, some of the studies were based on dataset images collected from internet sources.

3.7.2 NUMBER OF IMAGES ARE LIMITED

The datasets which were used in the selected studies have limited images for detecting and testing the thyroid nodules. Therefore, the reliability of the results might become uncertain if a large number of images will be available in the datasets.

3.7.3 PRIVATE DATASETS

Few researchers like [26], [27], [28], [29], [31], [34], [36],[38] had used non-public dataset and collect the images from internet due to which results cannot be replicated efficiently as proper dataset is not available.

3.7.4 LESION SIZE

During the literature review, it was observed that lesion size is very important for detecting and diagnosing the thyroid nodule. Lesion size of less than 1cm/10mm is quite difficult to diagnose or detect the nodules and hence the precision and accuracy of the diagnoses fell haphazardly. Lesion size above 1cm/10mm is suitable to detect thyroid nodules [7].

3.7.5 THYROID CLASSIFIERS

It was observed during the SLR study that pre-trained models which are based on deep learning often give compromising consequences for detecting the thyroid nodules [31-40],[46][48 -49], in terms of accuracy and effectiveness. After a thorough review of all the selected studies, it was noticed that deep learning models showed good results if 70% images of the datasets are utilized for training and 30% are used for testing. It was also seen that among the selected studies some have increased the percentage of the training group and got promised results. In short, we have deduced that the methods of deep learning are appropriate for diagnosing if optimum ratios for testing and trained groups are used.

4. CONCLUSION

Our study was based on a systematic literature review (SLR) that was conducted on the problem of thyroid diagnosis. In this study, we have discussed deep learning techniques including neural network (fully convoluted), ensemble, and handcrafted approaches to classify the thyroid nodules. It was revealed that the deep learning technology does not need to pass through any sort of complex procedures and pre-processing including resizing, cropping, and normalization of pixels in images. Moreover, this research work also suggests some improvements in the existing methods, which are required in avoiding the shortcomings or weaknesses in the areas where they might be indicated. It was observed that the handcrafted methods give better conclusions or results than the conventional methods of deep learning. Few studies which were based on handcrafted mechanism focus on the functionality of extraction process and segmentation. Besides this, image labeling was considered the most influential task in demographic imaging of relevant databases. The clinical implication is underpinning due to day-to-day advancement in p modules prepared and developed through deep learning technology. Furthermore, this advancement leads us towards artificial intelligence-based research technologies for medical diagnoses shortly. By applying the appropriate criteria for inclusion and exclusion here in our research the databases of only selected articles were considered which are presented in Table 2. Some of the databases and journals were not explored because our work did not fall under their scope. To avoid the language barriers, the research papers written in languages other than the English language have not been targeted therefore the papers which are published in the English language are selected as samples for this SLR study. For future research, it has been expected that more specialized deep learning algorithms will be designed and developed which may contribute efficiently in providing more accurate results in diagnosing disease through medical imaging.

CREDIT AUTHOR STATEMENT

Maheen Ilyas: Conceptualization, Methodology, Software implementation, Data curation, Writing-Original draft preparation. **Hassaan Malik**: Supervision. **Muhammad Adnan**: Methodology, Visualization, Writing-Original draft preparation. **Umair Bashir**: Conceptualization, Methodology, Software implementation, Data curation, Writing-Original draft preparation. **Wajahat Anwaar Bukhari**: Methodology, Software implementation, Writing- Reviewing and Editing. **Muhammad Imran Ali Khan**: Conceptualization, Methodology, Writing-Original draft preparation. **Adnan Ahmad**: Conceptualization, Methodology, Writing-Original draft preparation.

Compliance with Ethical Standards:

It is declare that all authors don’t have any conflict of interest. It is also declare that this article does not contain any studies with human participants or animals performed by any of the authors. Furthermore, informed consent was obtained from all individual participants included in the study.

REFRENCES

1. Davies, L., & Welch, H. G. (2006). Increasing incidence of thyroid cancer in the United States, 1973-2002. Jama, 295(18), 2164-2167.
2. Kim, T. Y., &Shong, Y. K. (2017). Active surveillance of papillary thyroid microcarcinoma: a mini-review from Korea. Endocrinology and Metabolism, 32(4), 399-406.
3. Ezzat, S., Sarti, D. A., Cain, D. R., &Braunstein, G. D. (1994). Thyroid incidentalomas: prevalence by palpation and ultrasonography. Archives of internal medicine, 154(16), 1838-1840.
4. Reiners, C., Wegscheider, K., Schicha, H., Theissen, P., Vaupel, R., Wrbitzky, R., &Schumm-Draeger, P. M. (2004). Prevalence of thyroid disorders in the working population of Germany: ultrasonography screening in 96,278 unselected employees. Thyroid, 14(11), 926-932.
5. Vickers, N. J. (2017). Animal Communication: When I’m Calling You, Will You Answer Too?. Current Biology, 27(14), R713-R715.
6. Werga, P., Wallin, G., Skoog, L., &Hamberger, B. (2000). Expanding role of fine-needle aspiration cytology in thyroid diagnosis and management. World journal of surgery, 24(8), 907-912.
7. Theoharis, C. G., Schofield, K. M., Hammers, L., Udelsman, R., &Chhieng, D. C. (2009). The Bethesda thyroid fine-needle aspiration classification system: year 1 at an academic institution. Thyroid, 19(11), 1215-1223.
8. McIver, B., Hay, I. D., Giuffrida, D. F., Dvorak, C. E., Grant, C. S., Thompson, G. B.,&Goellner, J. R. (2001). Anaplastic thyroid carcinoma: a 50-year experience at a single institution. Surgery, 130(6), 1028-1034.
9. Lin, S., Huang, H., Liu, X., Li, Q., Yang, A. K., Zhang, Q., & Chen, Y. (2014). Treatments for complications of tracheal sleeve resection for papillary thyroid carcinoma

with tracheal invasion. European Journal of Surgical Oncology (EJSO), 40(2), 176-181.
10. Cronan, J. J. (2008). Thyroid nodules: is it time to turn off the US machines? Radiology, 247(3), 602-604.
11. Zahir, S. T., Vakili, M., Ghaneei, A., &Heidari, F. (2016). Ultrasound assistance in differentiating malignant thyroid nodules from benign ones. Journal of Ayub Medical College Abbottabad, 28(4), 644-649.
12. Poole, D., Mackworth, A., & Goebel, R. (1998). Computational Intelligence.
13. Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.
14. Haugen, B. R., Sawka, A. M., Alexander, E. K., Bible, K. C., Caturegli, P. D., & Doherty, G. (2017). The ATA guidelines on management of thyroid nodules and differentiated thyroid cancer task force review and recommendation on the proposed renaming of eFVPTC without invasion to NIFTP. Thyroid, 27, 481-483.
15. Smith-Bindman, R., Lebda, P., Feldstein, V. A., Sellami, D., Goldstein, R. B., Brasic, N., &Kornak, J. (2013). Risk of thyroid cancer based on thyroid ultrasound imaging characteristics: results of a population-based study. JAMA internal medicine, 173(19), 1788-1795.
16. Brito, J. P., Gionfriddo, M. R., Al Nofal, A., Boehmer, K. R., Leppin, A. L., Reading, C., ... & Murad, M. H. (2014). The accuracy of thyroid nodule ultrasound to predict thyroid cancer: systematic review and meta-analysis. The Journal of Clinical Endocrinology & Metabolism, 99(4), 1253-1263.
17. Shin, J. H., Baek, J. H., Chung, J., Ha, E. J., Kim, J. H., Lee, Y. H.,& Choi, Y. J. (2016). Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology consensus statement and recommendations. Korean journal of radiology, 17(3), 370-395.
18. Park, C. S., Kim, S. H., Jung, S. L., Kang, B. J., Kim, J. Y., Choi, J. J., .&Jeong, S. H. (2010). Observer variability in the sonographic evaluation of thyroid nodules. Journal of Clinical Ultrasound, 38(6), 287-293.
19. Acharya, U. R., Swapna, G., Sree, S. V., Molinari, F., Gupta, S., Bardales, R. H.,& Suri, J. S. (2014). A review on ultrasound-based thyroid cancer tissue characterization and automated classification. Technology in cancer research & treatment, 13(4), 289-301.
20. Shankar, K., Lakshmanaprabu, S. K., Gupta, D., Maseleno, A., & De Albuquerque, V. H. C. (2020). Optimal feature-based multi-kernel SVM approach for thyroid disease classification. The journal of supercomputing, 76(2), 1128-1143.
21. Gupta, N., Jain, R., Gupta, D., Khanna, A., & Khamparia, A. (2020). Modified ant lion optimization algorithm for improved diagnosis of thyroid disease. In Cognitive Informatics and Soft Computing (pp. 599-610). Springer, Singapore.
22. Song, W., Li, S., Liu, J., Qin, H., Zhang, B., Zhang, S., &Hao, A. (2018). Multitask cascade convolution neural networks for automatic thyroid nodule detection and recognition. IEEE journal of biomedical and health informatics, 23(3), 1215-1224.
23. Chi, J., Walia, E., Babyn, P., Wang, J., Groot, G., &Eramian, M. (2017). Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. Journal of digital imaging, 30(4), 477-486.
24. B. Kitchenham, “Procedures for undertaking systematic reviews: joint technical report,” Dept. Comput. Sci., Keele Univ., Nat. ICT Australia, Keele, U.K., Tech. Rep. TR/SE-0401, 2004.
25. B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, “Systematic literature reviews in software engineering_A systematic literature review," Inf. Softw. Technol., vol. 51, no. 1, pp. 7_15, 2009.
26. Gharib, H., Papini, E., Paschke, R., Duick, D., Valcavi, R., Hegedüs, L., & Vitti, P. (2010). American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. Endocrine Practice, 16(Supplement 1), 1-43.
27. Santin, M., Brama, C., Théro, H., Ketheeswaran, E., El-Karoui, I., Bidault, F., ... & Blum, A. (2019). Detecting abnormal thyroid cartilages on CT using deep learning. Diagnostic and interventional imaging, 100(4), 251-257.
28. Zhou, H., Jin, Y., Dai, L., Zhang, M., Qiu, Y., Tian, J., & Zheng, J. (2020). Differential Diagnosis of Benign and Malignant Thyroid Nodules Using Deep Learning Radiomics of Thyroid Ultrasound Images. European Journal of Radiology, 108992.
29. Prochazka, A., Gulati, S., Holinka, S., & Smutek, D. (2019). Classification of thyroid nodules in ultrasound images using direction-independent features extracted by two-threshold binary decomposition. Technology in cancer research & treatment, 18, 1533033819830748.
30. Zhang, S., Du, H., Jin, Z., Zhu, Y., Zhang, Y., Xie, F., ... & Luo, Y. (2020). A Novel Interpretable Computer-Aided Diagnosis System of Thyroid Nodules on Ultrasound Based on Clinical Experience. IEEE Access, 8, 53223-53231.
31. Guan, Q., Wang, Y., Du, J., Qin, Y., Lu, H., Xiang, J., & Wang, F. (2019). Deep learning based classification of ultrasound images for thyroid nodules: a large scale of pilot study. Annals of Translational Medicine, 7(7).
32. Wang, L., Yang, S., Yang, S., Zhao, C., Tian, G., Gao, Y., ... & Lu, Y. (2019). Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network. World journal of surgical oncology, 17(1), 1-9.
33. Moon, J. H., & Steinhubl, S. R. (2019). Digital medicine in thyroidology: a new era of managing thyroid disease. Endocrinology and Metabolism, 34(2), 124-131.

34. Gitto, S., Grassi, G., De Angelis, C., Monaco, C. G., Sdao, S., Sardanelli, F., ... & Mauri, G. (2019). A computer-aided diagnosis system for the assessment and characterization of low-to-high suspicion thyroid nodules on ultrasound. La radiologia medica, 124(2), 118-125.
35. Jeong, E. Y., Kim, H. L., Ha, E. J., Park, S. Y., Cho, Y. J., & Han, M. (2019). Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. European radiology, 29(4), 1978-1985.
36. Kim, H. L., Ha, E. J., & Han, M. (2019). Real-world performance of computer-aided diagnosis system for thyroid nodules using ultrasonography. Ultrasound in medicine & biology, 45(10), 2672-2678.
37. Ko, S. Y., Lee, J. H., Yoon, J. H., Na, H., Hong, E., Han, K., ... & Lee, E. (2019). Deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. Head & Neck, 41(4), 885-891.
38. Liang, X. W., Cai, Y. Y., Yu, J. S., Liao, J. Y., & Chen, Z. Y. (2019). Update on thyroid ultrasound: a narrative review from diagnostic criteria to artificial intelligence techniques. Chinese Medical Journal, 132(16), 1974.
39. Liu, R., Li, H., Liang, F., Yao, L., Liu, J., Li, M., ... & Song, B. (2019). Diagnostic accuracy of different computer-aided diagnostic systems for malignant and benign thyroid nodules classification in ultrasound images: A systematic review and meta-analysis protocol. Medicine, 98(29).
40. Song, J., Chai, Y. J., Masuoka, H., Park, S. W., Kim, S. J., Choi, J. Y., ... & Yi, K. H. (2019). Ultrasound image analysis using deep learning algorithm for the diagnosis of thyroid nodules. Medicine, 98(15).
41. Zhao, W. J., Fu, L. R., Huang, Z. M., Zhu, J. Q., & Ma, B. Y. (2019). Effectiveness evaluation of computer-aided diagnosis system for the diagnosis of thyroid nodules on ultrasound: A systematic review and meta-analysis. Medicine, 98(32).
42. Li, X., Zhang, S., Zhang, Q., Wei, X., Pan, Y., Zhao, J., ... & Yang, F. (2019). Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. The Lancet Oncology, 20(2), 193-201.
43. Orloff, L. A. (2020). Thyroid Ultrasound: Machine Beats Humans at Detecting Malignant Nodules. Clinical Thyroidology, 32(2), 69-71.
44. Zhang, B., Tian, J., Pei, S., Chen, Y., He, X., Dong, Y., ... & Zhang, S. (2019). Machine learning–assisted system for thyroid nodule diagnosis. Thyroid, 29(6), 858-867.
45. Wildman-Tobriner, B., Buda, M., Hoang, J. K., Middleton, W. D., Thayer, D., Short, R. G., ... & Mazurowski, M. A. (2019). Using artificial intelligence to revise ACR TI-RADS risk stratification of thyroid nodules: diagnostic accuracy and utility. Radiology, 292(1), 112-119.
46. Xu, L., Gao, J., Wang, Q., Yin, J., Yu, P., Bai, B., ... & Wan, M. (2020). Computer-Aided Diagnosis Systems in Diagnosing Malignant Thyroid Nodules on Ultrasonography: A Systematic Review and Meta-Analysis. European Thyroid Journal, 9(4), 186-193.
47. Wang, Y., Wei, K., & Wan, P. (2018). A method of ultrasonic image recognition for thyroid papillary carcinoma based on deep convolution neural network. NeuroQuantology, 16(5).
48. Ouyang, F. S., Guo, B. L., Ouyang, L. Z., Liu, Z. W., Lin, S. J., Meng, W., ... & Yang, S. M. (2019). Comparison between linear and nonlinear machine-learning algorithms for the classification of thyroid nodules. European journal of radiology, 113, 251-257.
49. Sollini, M., Cozzi, L., Chiti, A., & Kirienko, M. (2018). Texture analysis and machine learning to characterize suspected thyroid nodules and differentiated thyroid cancer: Where do we stand?. European journal of radiology, 99, 1-8.
50. Song, W., Li, S., Liu, J., Qin, H., Zhang, B., Zhang, S., & Hao, A. (2018). Multitask cascade convolution neural networks for automatic thyroid nodule detection and recognition. IEEE journal of biomedical and health informatics, 23(3), 1215-1224.
51. Yoo, Y. J., Ha, E. J., Cho, Y. J., Kim, H. L., Han, M., & Kang, S. Y. (2018). Computer-aided diagnosis of thyroid nodules via ultrasonography: initial clinical experience. Korean journal of radiology, 19(4), 665-672.
52. Lauria, A. P., Maddaloni, E., Briganti, S. I., Beretta, G. A., Perrella, E., Taffon, C., ... & Crescenzi, A. (2018). Differences between ATA, AACE/ACE/AME and ACR TI-RADS ultrasound classifications performance in identifying cytological high-risk thyroid nodules. European journal of endocrinology, 178(6), 595-603.
53. Zhang, Y., Zhang, M. B., Luo, Y. K., Li, J., Wang, Z. L., & Tang, J. (2018). The value of peripheral enhancement pattern for diagnosing thyroid cancer using contrast-enhanced ultrasound. International journal of endocrinology, 2018.
54. Liu, C., Huang, Y., Ozolek, J. A., Hanna, M. G., Singh, R., & Rohde, G. K. (2018). SetSVM: an approach to set classification in nuclei-based cancer detection. IEEE Journal of Biomedical and Health Informatics, 23(1), 351-361.
55. Mauri, G., Pacella, C. M., Papini, E., Solbiati, L., Goldberg, S. N., Ahmed, M., & Sconfienza, L. M. (2019). Image-guided thyroid ablation: proposal for standardization of terminology and reporting criteria. Thyroid, 29(5), 611-618.
56. Caresio, C., Caballo, M., Deandrea, M., Garberoglio, R., Mormile, A., Rossetto, R., ... & Molinari, F. (2018). Quantitative analysis of thyroid tumors vascularity: A comparison between 3-D contrast-enhanced ultrasound and 3-D Power Doppler on benign and malignant thyroid nodules. Medical physics, 45(7), 3173-3184.
57. Sun, C., Zhang, Y., Chang, Q., Liu, T., Zhang, S., Wang, X., ... & Niu, L. (2020). Evaluation of a deep learning-based computer-aided diagnosis system for distinguishing

benign from malignant thyroid nodules in ultrasound images. Medical Physics, 47(9), 3952-3960.

58. Abdolali, F., Kapur, J., Jaremko, J. L., Noga, M., Hareendranathan, A. R., & Punithakumar, K. (2020). Automated thyroid nodule detection from ultrasound imaging using deep convolutional neural networks. Computers in Biology and Medicine, 122, 103871.
59. Li, Y., Chen, P., Li, Z., Su, H., Yang, L., & Zhong, D. (2020). Rule-based automatic diagnosis of thyroid nodules from intraoperative frozen sections using deep learning. Artificial Intelligence in Medicine, 108, 101918.
60. Li, L. R., Du, B., Liu, H. Q., & Chen, C. (2021). Artificial Intelligence for Personalized Medicine in Thyroid Cancer: Current Status and Future Perspectives. Frontiers in Oncology, 10, 3360.
61. Yoon, J., Lee, E., Koo, J. S., Yoon, J. H., Nam, K. H., Lee, J., ... & Kwak, J. Y. (2020). Artificial intelligence to predict the BRAFV600E mutation in patients with thyroid cancer. PloS one, 15(11), e0242806.
62. Han, M., Ha, E. J., & Park, J. H. (2021). Computer-Aided Diagnostic System for Thyroid Nodules on Ultrasonography: Diagnostic Performance Based on the Thyroid Imaging Reporting and Data System Classification and Dichotomous Outcomes. American Journal of Neuroradiology, 42(3), 559-565.
63. Guo, X., Zhao, H., & Tang, Z. (2020, April). An Improved Deep Learning Approach for Thyroid Nodule Diagnosis. In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI) (pp. 296-299). IEEE.
64. Chen, D., Zhang, J., & Li, W. (2018, October). Thyroid Nodule Classification Using Two Levels Attention-Based Bi-Directional LSTM with Ultrasound Reports. In 2018 9th Intenational Conference on Information Technology in Medcine and Education (ITME) (pp. 309-312). IEEE.
65. González, J. R., Conci, A., Moran, M. B. H., Araujo, A. S., Paes, A., Damião, C., & Fiirst, W. G. (2019, November). Analysis of Static and Dynamic Infrared Images for Thyroid Nodules Investigation. In 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA) (pp. 1-7). IEEE.
66. Zhang, H., Zhao, C., Guo, L., Li, X., Luo, Y., Lu, J., & Xu, H. (2019, October). Diagnosis of Thyroid Nodules in Ultrasound Images Using Two Combined Classification Modules. In 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) (pp. 1-5). IEEE.
67. Bhatti, M. H., Khan, J., Khan, M. U. G., Iqbal, R., Aloqaily, M., Jararweh, Y., & Gupta, B. (2019). Soft computing-based EEG classification by optimal feature selection and neural networks. IEEE Transactions on Industrial Informatics, 15(10), 5747-5754.
68. Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Dong, J. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 172(5), 1122-1131.
69. Long, E., Lin, H., Liu, Z., Wu, X., Wang, L., Jiang, J., ... & Li, J. (2017). An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nature biomedical engineering, 1(2), 1-8.
70. Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature medicine, 25(1), 65.
71. Acharya, U. R., Swapna, G., Sree, S. V., Molinari, F., Gupta, S., Bardales, R. H., ... & Suri, J. S. (2014). A review on ultrasound-based thyroid cancer tissue characterization and automated classification. Technology in cancer research & treatment, 13(4), 289-301.
72. Kitchenham, B. (2004). Procedures for performing systematic reviews. Keele, UK, Keele University 33(2004),1-26.
73. Giannoula, E., Iakovou, I., Katsikavelas, I., Antoniou, P., Raftopoulos, V., Chatzipavlidou, V., ... & Bamidis, P. (2020). A mobile app for thyroid cancer patients aiming to enhance their quality of life: protocol for a quasiexperimental interventional pilot study. *JMIR research protocols*, *9*(3), e13409.