A systematic review on artificial intelligence techniques for predicting thyroid diseases (#78828)

First submission

Guidance from your Editor

Please submit by **12 Feb 2023** for the benefit of the authors (and your token reward) .

Literature Review article
This is a Literature Review article, so the review criteria are slightly different.
Please write your review using the criteria outlined on the 'Structure and Criteria' page.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files
Download and review all files from the materials page.

7 Figure file(s)
2 Latex file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **STUDY DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Is the review of broad and cross-disciplinary interest and within the scope of the journal?
- Has the field been reviewed recently? If so, is there a good reason for this review (different point of view, accessible to a different audience, etc.)?
- Does the Introduction adequately introduce the subject and make it clear who the audience is/what the motivation is?

STUDY DESIGN

- Article content is within the Aims and Scope of the journal.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.
- Is the Survey Methodology consistent with a comprehensive, unbiased coverage of the subject? If not, what is missing?
- Are sources adequately cited? Quoted or paraphrased as appropriate?
- Is the review organized logically into coherent paragraphs/subsections?

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Meaningful replication encouraged where rationale & benefit to literature is clearly stated.
- Conclusions are well stated, linked to original research question & limited to
- Is there a well developed and supported argument that meets the goals set out in the Introduction?
- Does the Conclusion identify unresolved questions / gaps / future directions?

supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be*

improved upon before Acceptance.

A systematic review on artificial intelligence techniques for predicting thyroid diseases

Lerina Aversano [1] , **Mario Luca Bernardi** [1] , **Marta Cimitile** [2] , **Andrea Maiellaro** [1] , **Riccardo Pecori** Corresp. [3, 4]

[1] Department of Engineering, University of Sannio, Benevento, Italy
[2] UnitelmaSapienza University, Rome, Italy
[3] Faculty of Engineering, eCampus University, Novedrate (CO), Italia
[4] Institute of Materials for Electronics and Magnetism, National Research Council, Parma, Italy

Corresponding Author: Riccardo Pecori
Email address: riccardo.pecori@uniecampus.it

The use of artificial intelligence approaches in health-care systems has grown rapidly over the last few years. In this context, early detection of diseases is the most common area of application. In this scenario, thyroid diseases are a vital example of illnesses that can be effectively faced if discovered quite early. This work aims at systematically reviewing and analyzing the literature on various artificial intelligence- techniques applied to detect and identify of various diseases related to the thyroid gland. The contributions we reviewed are classified according to different viewpoints into a coherent and structured taxonomy in order to highlight pros and cons of the most recent research in the field. Aftera careful selection process, we selected and reviewed 74 articles, analyzing them according to three main research questions, i.e., which diseases of the thyroid gland are detected by different artificial intelligence techniques, which datasets are used to perform the aforementioned detection, and what types of data are used to perform the detection. The review demonstrates that the majority of the considered papers deal with supervised methods to detect hypo- and hyperthyroidism. The average accuracy of detection is high (96.92%), but the usage of private and outdated datasets with a majority of clinical data is preponderant. Finally, we discuss the outcomes of the systematic review, pointing out advantages, disadvantages and future developments in the application of artificial intelligence for thyroid diseases detection.

Deleted: is rapidly increasing in the last years

Deleted: Among the others, their use is growing for the early detection of some diseases in order to promptly plan an adequate patient care…

Commented [A1]: Before jumping to thyroid please provide sufficient context for why thyroid is important?

Deleted: proper example

Deleted: research landscape about

Deleted: related

Deleted: the detection

Deleted: identification

Commented [A2]: Avoid. This is for audience to decide.

Deleted: quite

Commented [A3]: Use simpler word.

Deleted: about

Deleted: to the

Deleted: detection of

A Systematic Review on Artificial Intelligence Techniques for Predicting Thyroid Diseases

Lerina Aversano[1], Mario Luca Bernardi[1], Marta Cimitile[2], Andrea Maiellaro[1], and Riccardo Pecori[3,4]

[1]**Dept. of Engineering, University of Sannio, Via Traiano 9, Benevento, 82100, Italy**
[2]**Unitelma Sapienza University, Piazza Sassari 4, Rome, 00161, Italy**
[3]**eCampus University, Via Isimbardi 10, Novedrate (CO), 22060, Italy**
[4]**IMEM-CNR, Parco Area delle Scienze 37/A, Parma, 43124, Italy**

Corresponding author:
Riccardo Pecori

Email address: riccardo.pecori@uniecampus.it

ABSTRACT

The use of artificial intelligence approaches in health-care systems is rapidly increasing in the last years. Among the others, their use is growing for the early detection of some diseases in order to promptly plan an adequate patient care. In this scenario, thyroid diseases are a proper example of illnesses that can be effectively faced if discovered quite early. This work aims at systematically reviewing and analyzing the research landscape about various artificial intelligence-related techniques applied to the detection and identification of various diseases related to the thyroid gland.

The contributions we reviewed are classified according to different viewpoints into a coherent and structured taxonomy in order to highlight pros and cons of the most recent research in the field.

After a careful selection process, we selected and reviewed 74 articles, analyzing them according to three main research questions, i.e., which diseases of the thyroid gland are detected by different artificial intelligence techniques, which datasets are used to perform the aforementioned detection, and what types of data are used to perform the detection.

The review demonstrates that the majority of the considered papers deal with supervised methods to detect hypo- and hyperthyroidism. The average accuracy of detection is quite high (96.92%), but the usage of private and outdated datasets with a majority of clinical data is preponderant.

Finally, we discuss about the outcomes of the systematic review, pointing out advantages, disadvantages and future developments in the application of artificial intelligence to the detection of thyroid diseases.

1 INTRODUCTION

Thyroid is an important human endocrine organ, positioned in the anterior part of the neck and secretes the hormone which regulates human metabolism. A thyroid disorder takes place whenever

Deleted: he thyroid
Deleted: in
Deleted: ing
Deleted: an
Deleted: able to regulate the

this organ produces either too many or too few hormones. Thyroid disorder is the most common disease in the endocrine field [Longbottom and Macnab, 2014], causing several ailments and in its more severe forms, also death. For this reason, it is very important to diagnose the disease at its early stages and take precautions to avoid the most dangerous conditions.

There are several approaches to diagnose thyroid disorder, such as the detection of the thyroid hormone, clinical test evaluation, imaging inspection, blood analysis, and tissue biopsy. These approaches require the drudgery of doctors and is not foolproof because the process of diagnosis of thyroid disorder from the laboratory analysis is complex and requires the doctors' extensive knowledge and experience. For this reason, in the last years, several research studies, have used Artificial Intelligence (AI) techniques to predict various thyroid diseases [Ma et al., 2019; Kwak and Hui, 2019]. Indeed, AI extensively used to solve problems in the health field [Kwak and Hui, 2019; Aversano et al., 2021] and has provided good accuracy. This is further spurred by the increasing computational power of computers, which allow running the most complex and time-consuming AI algorithms. Despite, the increasing number of studies about AI to predict thyroid diseases, a broader discussion and comparison of the used AI approaches in the given context is lacking.

The rationale of this proposed systematic review is to investigate the use of AI approaches for the early detection of thyroid dysfunctions, and provide mapping of the used data types and the available datasets. The audience for this review includes computer scientists, bio-informatics specialists, data analysts as well as medical doctors and endocrinologists in particular.

The main objectives of this systematic review are the following:

- summarizing the most recent AI solutions linked to the early prediction of thyroid diseases (Research Question RQ1);
- identifying the used datasets to apply AI solutions for the early prediction of thyroid diseases (Research Question RQ2);
- summarizing the most used data types to detect thyroid diseases using AI techniques (Research Question RQ3).

The rest of the article has the following structure. In Section 2 background concepts useful to understand the proposed investigation are reported. In Section 3, we summarize, from a critical point of view, some recent reviews obtained in the analysis and filtering process. In Section 4, the research method adopted in this systematic review is described, together with the research questions, the databases used, the keywords, and the filtering as well as the inclusion/exclusion criteria. In Section 5, we summarize and classify the considered articles, in a way suitable to answer the research questions. Finally, Section 6 reports a discussion of the obtained findings, also highlighting the current research gaps, while Section 7 concludes the article with an overview of the systematic review and of the obtained results.

Deleted: one of the largest diseases
Deleted: evaluate the
Deleted: take to
Commented [A4]: Unclear.
Commented [A5]: Simpler word.
Deleted: and can
be subject to some mistakes since
Formatted: Space Before: 6.4 pt
Deleted: quite
Deleted: using
Deleted: specifically
Deleted: have been carried out
Deleted: is more
and more
Deleted: with a
Deleted: , also thanks to
Deleted: However
Deleted: is not followed by
Deleted: also with respect to the various thyroid diseases themselves…
Deleted: specifically
Deleted: usages
Deleted: providing
Deleted: also a
Deleted: possible
Deleted: encompasses
Deleted: Indeed, the
Deleted: for answering to the need for
Deleted: some
Deleted: the comprehension of the following
Deleted: , described in details in
Section 4
Deleted: Herein
Formatted: Space Before: 6.4 pt
Deleted: used

2 BACKGROUND

The primary function of the thyroid gland is the production of the triiodothyronine (T3) and thyroxine (T4) hormones. These hormones travel through the body and help in the regulation of the metabolism, while also aiding brain development, digestive function, muscle control, and mood balancing. Autoimmune diseases and nutrient deficiencies are the principal causes of thyroid complications [Monaco, 2003]. Thyroid dysfunction is rather common in the general population, and mild or sub-clinical forms can be present in more than 10% of individuals older than 80 years.

The diagnosis of abnormal thyroid hormone concentrations in people older than 60 years poses a challenge, as the clinical presentation of thyroid dysfunction is usually nonspecific, and aging is associated with a number of physiological changes that can affect thyroid function test results. There are different kinds of thyroid dysfunction, namely, they are goiter, hyperthyroidism, hypothyroidism, malignant thyroid nodules, thyroiditis, etc. They are briefly summarized in the following [Monaco, 2003]:

- Goiter is a noncancerous enlargement of the thyroid gland. The most common cause of goiter worldwide is iodine deficiency in the diet;
- Hyperthyroidism occurs when the thyroid gland is overactive. It produces too much of its hormone;
- Hypothyroidism is the opposite of hyperthyroidism. The thyroid gland is underactive, and it cannot produce enough of its hormones;
- Thyroid nodules are growths that form on or in the thyroid gland. The nodules can be solid or fluid-filled, most are benign, but they can also be cancerous in a small percentage of cases;
- Thyroiditis can be considered a swelling of the thyroid.

Another pathology is Euthyroid, a normal thyroid hormonal functional state, but involved in initial structural changes such as goiter, cold nodule, multiple nodule goiter (MNG) and cancer (Grave's Disease and so on).

In the rest of the paper, we will specifically focus on the following general and more frequent thyroid-related diseases: hypo and hyperthyroidism, thyroid cancer, and euthyroid sick state. Other thyroid-related diseases will be considered, but in a general thyroid disease category.

3 RELATED WORK

In this section, we summarize the existing reviews and surveys on the studied topic and we highlight the differences, compared with our proposed literature review.

Garg and Mago [2021] only focus on the role of machine learning in medical research in general, Hasanzad et al. [2022] focus on endocrinology in general, [Kumar et al., 2022b] is a generic review on the application of artificial intelligence to the identification of diseases, Kumar et al. [2022a] concentrate on artificial intelligence and general cancer prediction, [Liu et al., 2021a] is a review about the general

Deleted: The thyroid gland, placed at the base of the neck, has as its…

Deleted: but

Deleted: they

Deleted: the

Deleted: the

Deleted: the

Deleted: the balancing of the mood

Commented [A6]: Please use smaller paragraphs to enhance readability.

Commented [A7]: The importance of talking about these is not immediately clear.

application of artificial intelligence in medicine, while [Wilson et al., 2022] is a survey on the application of artificial intelligence in otolaryngology in general.

In [Parmar and Mehta, 2020], the authors discuss different kinds of dysfunctions which affect the thyroid and point out the main methods and processes used currently for detecting these dysfunctions, but only from a clinical perspective. Moreover, the authors illustrate the computed supported detection techniques distinguishing them on the basis of the used input modes. The paper also outlines the main strengths and open research problems to be addressed in this research area. A set of parameters, such as, …… is used to compare the discussed techniques.

Deleted: report a survey discussing the
Deleted: affecting
Deleted: pointing
Deleted: currently
Deleted: prospective
Commented [A8]: unclear

The study in [Chen et al., 2020] proposes a review and categorization of thyroid gland segmentation and thyroid nodule segmentation methods according to the theoretical bases of segmentation methods. In particular, the review compared 28 representative papers selected in the literature. The most common methods for thyroid gland are based on machine and deep learning methods. Moreover, the study found out that big data for training provide segmentation performance and robustness. However, deep learning models usually require large training dataset and imply long training time. For thyroid nodule segmentation, the most common adopted methods are contour and shape based methods, that lead to satisfactory performance results. Nevertheless, they are often tested on small datasets.

Deleted: entailed a comprehensive analysis and comparison of…
Deleted: largely adopted segmentation
Deleted: it emerged from this study that the training of massive data…
Deleted: makes¶
114 these models have better
Deleted: satisfying

In [Abdolali et al., 2020], the authors provide a systematic review of artificial intelligence application focused on thyroid cancer diagnosis. The review considered and classify more than 50 papers discussing approaches for thyroid cancer detection exploiting AI algorithms. The paper also proposes future trends and challenges in the field and perspectives of computer-aided analysis to improve the efficiency of future methods for thyroid cancer diagnosis.

Finally, [Razia et al., 2020] is not a systematic review, but only a general survey of various machine learning techniques in medicine, Mendoza and Hernandez [2021] provide a review in the limited room of a conference paper made of six pages and they do not provide any keywords or research questions, thus they do not adhere to Barbara Kitchenham's guidelines [Kitchenham, 2004] for a systematic review either. Other recent reviews considering AI in the context of improving thyroid health, focus only on images, such as [Bini et al., 2021] and [Sharifi et al., 2021], and do not consider clinical data at all.

Commented [A9]: Please discuss the implications of such an approach and tie it to your aims.

Differently from all the aforementioned papers, the proposed systematic literature review aims to perform a comprehensive investigation on the use of artificial intelligence approaches for the diagnosis and classification of the main thyroid dysfunctions, providing the mapping of the considered approaches with the used data types and the available datasets. We do not concentrate only on images or only on clinical data, we do not focus only on specific thyroid dysfunctions, we do not regard works tackling thyroid diseases only as an example of application of machine learning techniques; moreover, we carefully adhere to Barbara Kitchenham's guidelines for a systematic review, clearly pointing out the investigated research questions, the formal analysis and filtering process we have followed, as well as the final results.

Deleted: specifically
Deleted: also
Commented [A10]: It is better to write in an affirmative tone. For e.g. We consider data modalities like image, clinical data, include all thyroid dysfunctions….

4 RESEARCH METHOD

In order to properly conduct the literature review, we adopted the guidelines proposed by Kitchenham [2004]. The phases we sequentially performed are listed in the following:

- definition of the research questions;
- extraction of the relevant keywords, from the research questions, that will be used for formulating the queries;
- database selection, in order to identify the scientific databases as sources for performing the initial search;
- definition of some initial filtering criteria, such as the time interval of the search, the venue, and quality of the searched results, etc.
- skimming of abstracts and papers to exclude irrelevant articles and possible duplicates;
- definition of eligibility criteria and their application during the full reading of the surveyed papers;
- full reading and analysis of the remaining articles considering the defined research questions.

Commented [A11]: If this is a standard practice according to Kitchenham, then you may omit these details here.

4.1 Research Questions and relevant keywords

The research questions we used to investigate the application of AI techniques to the classification of thyroid diseases are the following:

- RQ1: What are the main AI techniques used for the classification and identification of the most relevant thyroid diseases?
- RQ2: What datasets about thyroid diseases are used in the considered AI solutions?
- RQ3: What data types are used to detect and classify thyroid diseases using the considered AI techniques?

The first research question aims to discriminate the various thyroid diseases tackled by different AI-based classification approaches as well as their performance metrics. In analyzing the papers, we considered the following macro categories for the AI-based identification and classification approaches [Witten, 2011]:

Deleted: wants

1. Probabilistic Approaches, which classify or group samples according to a certain probability distribution function;
2. Kernel-based approaches, which perform pattern analysis by transforming linearly inseparable data to linearly separable ones;
3. Techniques using Decision Trees, a model often used in operational research exploiting a tree-like structure of decisions on features, each representing a node of the tree itself;

Deleted: p

Deleted: a

Deleted: k

t

Deleted: t

Deleted: d

Deleted: t

4. Rule-based approaches, which determines interpretable classification strategies by means of relational rules made of certain antecedents and a certain consequent;

5. Neural Networks, which are models inspired by human brain made of layers of artificial perceptrons, connected in different ways, and useful for both classification and description tasks;

6. cluster-based approaches, based on the division of the group of samples according to a certain similarity metric;

7. ensemble approaches, which combine different ML approaches and usually provides an output according to a certain strategy (e.g., majority voting);

8. other techniques.

Conversely, for the thyroid disease we focused on the following classification:

1. hypothyroidism, a situation of underactive thyroid gland when it does not produce enough of its crucial hormones;

2. hyperthyroidism, which is the opposite of hypothyroidism, is the situation when the thyroid gland produces too many of its hormones;

3. euthyroid disease, which can cause abnormal findings on thyroid function tests occurring in absence of any thyroidal illness;

4. cancer and malignant nodules;

5. other thyroid-related issues or general dysfunctions.

With reference to RQ2, we have analyzed the characteristics of the datasets used in the research studies, as well as their nature (private or public); whereas as regards RQ3 we have investigated whether the datasets were composed of images or clinical data, as well as the detailed list of considered features.

We have converted the research questions into proper queries, used to search certain databases described in the following. The used queries (Q) are the following, wherein query 2 is used for answering both research question 2 and 3:

1. Q1: (“artificial intelligence” OR “machine learning” OR “deep learning” OR “neural network” OR “neural networks”) AND (“thyroid disease” OR “thyroid diseases”);

2. Q2: (“artificial intelligence” OR “machine learning” OR “deep learning” OR “neural network” OR “neural networks”) AND (“thyroid disease” OR “thyroid diseases”) AND “dataset”;

Deleted: r

Deleted: n

Deleted: n

Commented [A12]: Please capitalize these names.

4.2 Searched databases

The selected papers were found in the following four main databases:

1. the database of the Institute of Electrical and Electronics Engineers (IEEE)[1], also called IEEEXplore, which contains technical articles in electrical engineering, electronics, computer science, and other related fields;
2. the Elsevier database[2], also called ScienceDirect, which permits to access to journals and technical and science articles in several scientific areas and fields;
3. the Springer database[3], also called SpringerLink, which permits one to access publications by the Springer Nature editorial group;
4. the database maintained by the Association for Computing Machinery[4], also called ACM Digital Library.

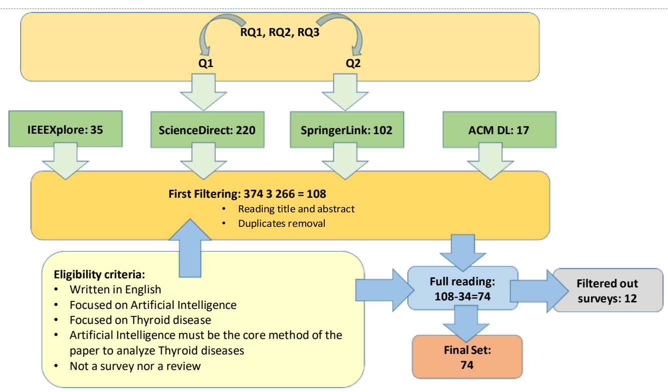


Figure 1. Details of the selection procedure for the analyzed papers.

Commented [A13]: In the part where, Q1 and Q2 is written, is this only connected to ScienceDirect and Springer?

2. Please provide a number in brackets for Duplicates removal () and Reading title and abstract ()

4.3 Search process and filtering criteria

We have performed the search and filtering process represented in Figure 1, wherein the inclusion (*IC*) and exclusion criteria (*EC*) presented in Table 1 were applied.

[1]https://ieeexplore.ieee.org/Xplore/home.jsp
[2]https://www.sciencedirect.com/
[3]https://link.springer.com/
[4]https://dl.acm.org/

Table 1. Criteria for including or excluding paper in the performed research process.

Acronym	Description of the criterium
Inclusion Criteria	
IC_1	Studies published in the range 2016-2022
IC_2	Studies written in English
IC_3	Studies should use AI techniques for analyzing any thyroid disease
Exclusion Criteria	
EC_1	The article is a survey or review
EC_2	The research does not consider artificial intelligence
EC_3	The paper is not specifically focused on the thyroid

The selected papers fall into January 2016 - July 2022 (IC_1). We have chosen this range in order to analyze only the most recent studies, dating back to at most 6 years ago. After applying the queries to the aforementioned databases we got a total of 374 papers, 35 from IEEEXplore, 220 from ScienceDirect, 102 from SpringerLink and 17 from the ACM Digital Library.

Then, we skimmed titles and abstracts in order to remove possible duplicate titles, and articles that i) do not deal mainly with AI-based methods (EC_2), ii) do not consider the thyroid gland and its diseases as the main focus (EC_3).

After this phase, a total of 108 papers remained. Finally, these remaining articles were read fully for both extracting useful statistical information presented in the following and for filtering them further on the base of the following selection criteria:

- written in English (IC_2);
- specifically focused on thyroid (EC_3);
- an AI-based technique must be the core method of the paper to identify or classify thyroid diseases (IC_3);
- not a survey nor a review (EC_1).

This final filtering step resulted in a total of 74 papers to analyze after removal of 12 recent surveys, already discussed in Section 3.

Figure 2 presents the distribution over the years of the 74 papers selected for the systematic review. As one can see, there is a constant growth of the research on the topic over the recent years with the obvious exception of 2022, given that it can only consider half of the year. The trend of the graph demonstrates an increasing interest in the usage of AI-based methods to classify and identify thyroid-related dysfunctions, thus corroborating our initial idea to research on this hot topic.

5 RESULTS

In this section, we discuss the main results and outcomes of the systematic literature review, by following the research questions described in Section 4.

Commented [A14]: The year range 2016-2022

Deleted: temporal interval

Deleted: coming

Deleted: have

Deleted: duplicated

Deleted: papers

Deleted: survived

Deleted: surviving

Deleted: completely

Deleted: has allowed us to keep

Deleted: and to extract

Deleted: further

Deleted: that survived our sieving procedure

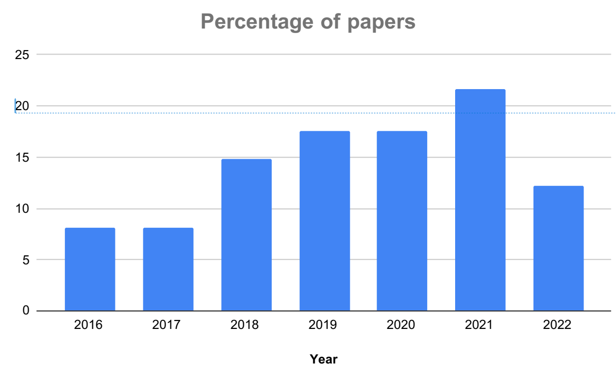


Figure 2. Percentage of the considered papers over the years.

5.1 RQ1: What are the main AI techniques used for the classification and identification of the most relevant thyroid diseases?

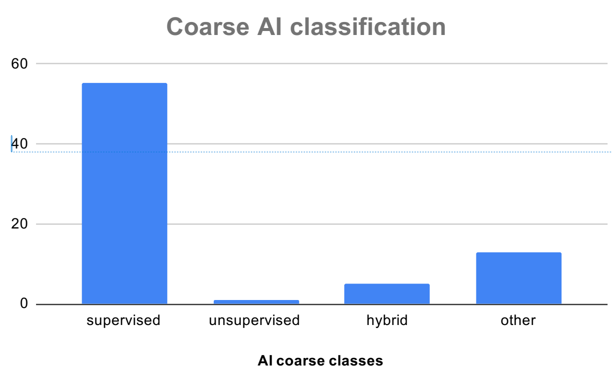


Figure 3. Percentage of the survived papers according to a coarse classification of the AI techniques.

This research question regarded the identification of the various AI techniques, as well as the main tackled thyroid diseases, used in the 74 papers identified after the filtering process.

Figure 3 summarizes the percentage of the identified papers as regards a classification of the AI techniques. This classification entails four main classes, namely, supervised, unsupervised, hybrid,

Commented [A15]: y-axis title missing. Suggested title for figure: Distribution of papers organized by publication year.

Commented [A16]: y-axis title missing. Please improve the title: AI methods used for thyroid prediction.

Deleted: described in the¶
240 previous section.

Formatted: Body Text, Space Before: 0 pt

Deleted: coarse

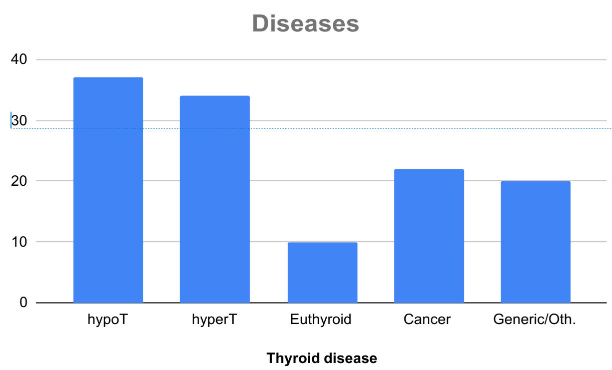
and other methods. The first ones regard techniques wherein the classification is performed by exploiting labels that are already known, the second ones try to find out a classification with no prior knowledge of any label, the third ones exploit a mix of both of the aforementioned, while the last ones use approaches which cannot be framed in the previous categories.

From the figure, it can be noted that the most of the papers (74.32%) entails supervised techniques, while unsupervised approaches are a very small fraction (1.35%). Hybrid methods represent 6.76%, while other methods are quite relevant, reaching a percentage equal to 17.57%.

Commented [A17]: Please specifically try to elaborate on how the techniques were combined.

In Figure 4, there is a percentage classification of the considered AI techniques on the basis of the four main thyroid disease categories we have considered. In this case, a study may face more than one disease, thus the percentage is not computed over 74 papers but over 123 tackled diseases. As one can see, the most studied diseases using AI techniques are hypothyroidism and hyperthyroidism (30.08% and 27.64% respectively), while euthyroid disease and the various types of thyroid cancer (or malignant nodules) reach a similar percentage of 8.13% and 17.89%, respectively. Finally, the generic thyroid dysfunctions and other thyroid-related issues encompass 16.26%.

Commented [A18]: Please explain how this number is 123.



Commented [A19]: y-axis title is missing.

Figure 4. Classification of the considered papers on the basis of the tackled thyroid disease.

In Table 3, we distribute the surveyed papers according to both the AI techniques and the just mentioned tackled thyroid diseases. The table also groups the various AI techniques according to the raw taxonomy described in Subsection 4.1.

In Figure 5, we show a percentage classification of the considered AI taxonomy. Also in this case, a study may exploit more than one technique, thus the percentage is not computed over 74 papers, but over 126 used techniques, without considering repetitions in each AI group. As one can see, the most used techniques to face thyroid diseases are neural networks, building up more than one third of the overall AI-based methods (34.13%), followed by Decision Trees (15.87%), kernel-based methods (15.08%),

Commented [A20]: Please clarify the basis of this figure.

Table 2. Statistics of the metrics of the surveyed AI techniques for thyroid disease classification.

	Accuracy	Precision	Recall	F1-score
Average	96.92%	94.83%	89.50%	89.50%
Maximum	100.00%	100.00%	100.00%	99.00%
Minimum	93.00%	85.00%	69.00%	74.00%

probabilistic techniques (9.52%), ensemble techniques (7.94%), and cluster-based methods (7.14%). The other methods represent a fraction smaller than 7%.

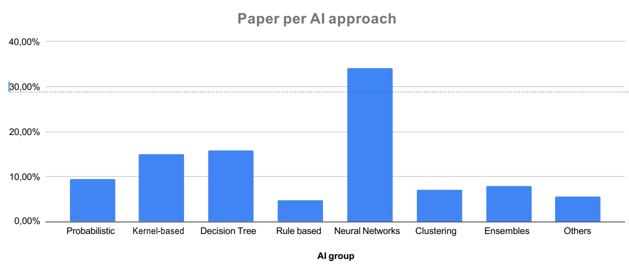


Figure 5. Percentage of the various AI groups.

Finally, Table 2 presents the average, maximum, and minimum values of the main metrics for classification (accuracy, precision, recall, and F1-score) obtained by the considered AI techniques in the surveyed papers.

Commented [A21]: Suggestion: Kindly arrange in descending order for better analysis of trend. Y-axis title is missing.

Table 3. Cross classification of the considered papers: according to the tackled thyroid disease and according to the various AI techniques, grouped also per macro-category.

		Thyroid diseases				
AI group	AI technique	Hypothyroidism	Hyperthyroidism	Euthyroid disease	Cancer	Generic/others
Probabilistic approaches	Naive Bayes	Rao and Renuka [2020]Pasha and Mohamed [2020]Houari et al. [2016] Chandel et al. [2016]Duggal and Shukla [2020]Peya et al. [2021] Riajuliislam et al. [2021]Juneja [2022]	Rao and Renuka [2020]Pasha and Mohamed [2020]Houari et al. [2016] Chandel et al. [2016]Duggal and Shukla [2020]Peya et al. [2021] Juneja [2022]	Duggal and Shukla [2020]	Qin et al. [2021]	Juneja [2022] Kishor and Chakraborty [2021]
	Log regression	Pasha and Mohamed [2020] Riajuliislam et al. [2021]	Pasha and Mohamed [2020]		Qin et al. [2021]	Raisinghani et al. [2019]Chaubey et al. [2021]
	TSR	Chandio et al. [2016]	Chandio et al. [2016]	Chandio et al. [2016]		
Kernel-based approaches	SVM	Duggal and Shukla [2020]Shahid et al. [2019]Ahmed and Soomrani [2016] Tyagi et al. [2018]Shen et al. [2016] Li et al. [2019b] Chandel et al. [2016]Zarin Mousavi et al. [2020]Pavya and Srinivasan [2017] Riajuliislam et al. [2021]	Duggal and Shukla [2020]Shahid et al. [2019]Ahmed and Soomrani [2016] Tyagi et al. [2018]Shen et al. [2016]Li et al. [2019b] Chandel et al. [2016]Pavya and Srinivasan [2017]	Ahmed and Soomrani [2016]Shen et al. [2016]Li et al. [2019b] Duggal and Shukla [2020]	Raghavendra et al. [2018]Prochazka et al. [2019] Raghavendra et al. [2017]Qin et al. [2021]Shen et al. [2021]	Tyagi et al. [2018]Kaur et al. [2019] Khudhair Abbas [2021]Kishor and Chakraborty [2021]
	multi-kernel SVM	Shankar et al. [2018]-Kumar [2020]	Shankar et al. [2018]-Kumar [2020]	Kumar [2020]		Shankar et al. [2018]
Decision Tree approaches	Decision Tree	Rao and Renuka [2020]Sidiq and Mutahar Aaqib [2019]Tyagi et al. [2018] Hayashi [2017]Peya et al. [2021]Riajuliislam et al. [2021] Juneja [2022]	Rao and Renuka [2020]Sidiq and Mutahar Aaqib [2019]Tyagi et al. [2018] Hayashi [2017]Peya et al. [2021]Juneja [2022]		Hao et al. [2018]	Tyagi et al. [2018]Raisinghani et al. [2019]Kaur et al. [2019] Chaubey et al. [2021]Juneja [2022]Kishor and Chakraborty [2021]
	DS		Jha et al. [2018]			
	Random Forest	Sidiq and Mutahar Aaqib [2019]Shahid et al. [2019]Duggal and Shukla [2020] Riajuliislam et al. [2021]Juneja [2022]	Sidiq and Mutahar Aaqib [2019]Shahid et al. [2019]Imbus et al. [2017] Duggal and Shukla [2020]Juneja [2022]	Duggal and Shukla [2020]	Prochazka et al. [2019] Qin et al. [2021]	Pan et al. [2016]Prochazka et al. [2019]Raisinghani et al. [2019] Kaur et al. [2019]Juneja [2022]
	ID3	Zarin Mousavi et al. [2020]				
	CHAID	Zarin Mousavi et al. [2020]				
Rule-based approaches	Re-RX	Hayashi [2017]	Hayashi [2017]			
	Case-based reasoning	Bentaiba-Lagrid et al. [2020]	Bentaiba-Lagrid et al. [2020]			
	RBS		Imbus et al. [2017]			
	Fuzzy RBS	Asaad Sajadi et al. [2019]Kumari and Sharma [2019]	Kumari and Sharma [2019]			
Neural Network	ANN	Sidiq and Mutahar Aaqib [2019]Mahurkar and Gaikwad [2017]Tyagi et al. [2018] Houari et al. [2016]Vivar et al. [2020] Zarin Mousavi et al. [2020]	Sidiq and Mutahar Aaqib [2019]Mahurkar and Gaikwad [2017] Tyagi et al. [2018]Houari et al. [2016]Vivar et al. [2020]		Ahmed et al. [2022]Cordes et al. [2021]Jin et al. [2021]	Tyagi et al. [2018]Raisinghani et al. [2019]Khudhair Abbas [2021] Kaur et al. [2019]Santos et al. [2019]Kishor and Chakraborty [2021]
	BLSTM-LSTM	Yue et al. [2020]	Yue et al. [2020]Lu et al. [2020]			Chai [2020]
	CNN	Yue et al. [2020] Ananthi et al. [2022] Khan [2021]	Yue et al. [2020] Ananthi et al. [2022]		Yin et al. [2019]Yi et al. [2017]Lyu and Haque [2018] Li et al. [2019a]Moran et al. [2018]Ananthi et al. [2022] Chu et al. [2021] Liu et al. [2021b]Santillan et al. [2021] Song et al. [2022]	Poudel et al. [2019]Guo and Du [2019] Ananthi et al. [2022] Pi et al. [2022]Yang et al. [2021]
	GAN	Zhang et al. [2020]	Zhang et al. [2020]		Shi et al. [2020]Zhao et al. [2022]	
	Capsule Network				Ai et al. [2022]	
	RBF NN	Juneja [2022]	Juneja [2022]			Juneja [2022]
	Autoencoder	Saktheeswari and Balasubramanian [2021]	Saktheeswari and Balasubramanian [2021]		Saktheeswari and Balasubramanian [2021]	
	RNN				Santillan et al. [2021]	
	MLP BackPropagation	Yue et al. [2020]Zarin Mousavi et al. [2020]Pavya and Srinivasan [2017] Hosseinzadeh et al. [2021]Juneja [2022]	Yue et al. [2020]Pavya and Srinivasan [2017]Hosseinzadeh et al. [2021] Juneja [2022]		Qin et al. [2021]	Juneja [2022]
Cluster-based approaches	KNN	Shahid et al. [2019]Tyagi et al. [2018]Pasha and Mohamed [2020] Houari et al. [2016]Chandel et al. [2016]Pasha and Mohamed [2020] Peya et al. [2021]	Shahid et al. [2019]Tyagi et al. [2018]Pasha and Mohamed [2020] Houari et al. [2016]Chandel et al. [2016]Peya et al. [2021]		Qin et al. [2021]	Tyagi et al. [2018]Kaur et al. [2019] Chaubey et al. [2021] Kishor and Chakraborty [2021]
Ensembles	AdaBoost	Zarin Mousavi et al. [2020]Yadav and Pal [2022]Priyadharsini and Sasikala [2022]	Priyadharsini and Sasikala [2022]			Kishor and Chakraborty [2021]
	Stacking	Yadav and Pal [2022] Sidiq and Mutahar Aaqib [2019]	Sidiq and Mutahar Aaqib [2019]Jha et al. [2018]			
	Bagging	Zarin Mousavi et al. [2020]Yadav and Pal [2022] Priyadharsini and Sasikala [2022]	Priyadharsini and Sasikala [2022]		Qin et al. [2021]	
	Vote ensemble	Sidiq and Mutahar Aaqib [2019]Yadav and Pal [2022]	Sidiq and Mutahar Aaqib [2019]Jha et al. [2018]		Yin et al. [2019]	Pan et al. [2016]
	Dynamic ensemble	Alam et al. [2020]	Alam et al. [2020]	Alam et al. [2020]		
Other Learning approaches	Semi supervised Learning					Devi and Anita [2018]
	WELM					Priya and Manavalan [2018]
	ELM	Pavya and Srinivasan [2017]Ma et al. [2018] Juneja [2022]	Pavya and Srinivasan [2017]Ma et al. [2018] Juneja [2022]			Juneja [2022]
	ATOVIC	Baccour [2018]	Baccour [2018]			

Commented [A22]: Kindly put abbreviations of methods in footnotes.

For column on Cancer, do you mean Thyroid Cancer.

You can shrink the width of column titles "Euthyroid disease".

In the following subsections, we briefly summarize each surveyed paper according to the AI group presented in Table 3.

5.1.1 Probabilistic approaches

In [Rao and Renuka, 2020] Naive Bayes and a Decision Tree, built using the ID3 algorithm, are used to perform a binary prediction about whether the patient is affected by hypo- or hyperthyroidism.

The authors of the study in [Pasha and Mohamed, 2020] perform feature selection on the UCI dataset for thyroid disease by exploiting both a Random Forest-based method and a Gain Ratio technique. Finally, the prediction is performed by comparing different machine learning techniques, namely K-Nearest-Neighbor, Logistic Regression, and Naive Bayes. Similarly, in [Houari et al., 2016] the authors try to reduce redundant dimensions by exploiting Copulas and LU-decomposition techniques and test their methods also on the UCI thyroid dataset for detecting both hypo- and hyperthyroidism. For the evaluation of the data reduction techniques, the authors employed Naive Bayes, besides Artificial Neural Networks (ANN) and k-nearest neighbors (k-NN) as learning algorithms. Finally, the contribution in [Chandel et al., 2016] compares different ML techniques in detecting thyroid-related diseases (i.e., hypo- and hyperthyroidism), with a particular focus on Naive Bayes, K-Nearest Neighbor, and Support Vector Machine. Juneja [2022] presents a fuzzy adaptive feature filtration and expansion-based model to generate a novel feature set related to thyroid. The obtained feature set is then analyzed through Extreme Learning Machine classifier whose performance is compared with Na¨ıve Bayes, Decision Tree, Multilayer Perceptron, and Radial Basis Function networks. Kishor and Chakraborty [2021] compare seven machine learning classifiers such as decision tree, support vector machine, Na¨ıve Bayes, adaptive boosting, Random Forest, artificial neural networks, and K-nearest neighbor in order to predict fatal diseases about thyroid, finding out that random forest is the best performing algorithm. Qin et al. [2021] try to study papillary thyroid cancer through the analysis of magnetic resonance imaging radiomics using eight classifiers including logistic regression, bagging, random forests, extremely randomized trees, support vector machines, Na¨ıve Bayes, multilayer perception, and K-nearest neighbors. Some of the models succeeded into reaching a performance of correct classification higher than 95%.

In [Peya et al., 2021] a thyroid diseases prediction model is proposed through three machine learning classification algorithms, i.e., K-Nearest Neighbor, Naive Bayes, and Decision Trees. Using the thyroid data of the UCI machine learning repository and a 10-fold cross-validation, the performance of the three algorithms is tested and the decision tree resulted the most accurate, with a 99.7% of accuracy.

Riajuliislam et al. [2021] try to predict early stage hypothyroidism by applying three different feature selection procedures, namely RFE, UFS, and PCA, together with different classification algorithms, i.e., support vector machine, decision tree, random forest, logistic regression, and Naive Bayes. RFE applied to the UCI thyroid dataset allowed the authors to achieve a constant 99% accuracy value over the various classification algorithms.

In Raisinghani et al. [2019], the authors compare different machine learning approaches, namely, Logistic regression, Decision trees, Random forest, Support vector machine, to develop predictive models

Commented [A23]: Please use smaller, more readable paragraphs.

Commented [A24]: Please introduce full-forms at first use.

to detect a generic thyroid disease.

The contribution in Chandio et al. [2016] exploits Time Series Regression to create an intelligent system for thyroid disease visualization for a careful surveillance of the thyroid disease, with a particular focus on hypothyroidism, hyperthyroidism, and euthyroid disease.

Chaubey et al. [2021] compare logistic regression, decision tree, and kNN on the UC Irvin knowledge discovery database for thyroid diseases detection, obtaining the best result of accuracy from the kNN classifier (96.87%).

5.1.2 Kernel-based approaches

Duggal and Shukla [2020] perform feature selection and extraction before applying Naive Bayes, Support Vector Machine, and Random Forest to identify of hypothyroidism, hyperthyroidism, and euthyroidism.

Deleted: the identification

The authors of Shahid et al. [2019] compare Random Forest, Support Vector Machine, and K-Nearest Neighbours, on the UCI thyroid dataset, to discover the best performing algorithm, resulted to be Random Forest, in detecting hypo- and hyperthyroidism.

The contribution in [Ahmed and Soomrani, 2016] provides a framework, named Thyroid Disease Types Diagnostics (TDTD), aiming at making a diagnosis of various thyroid diseases in a very structured and transparent manner and exploiting binary and multi-SVM algorithms, as well as Bayesian isotonic regression for missing values.

In [Tyagi et al., 2018], the authors present again the results in detecting hypo- and hyperthyroidism (by using the UCI dataset) by means of different machine learning techniques, such as decision trees, artificial neural networks, support vector machines, and k-nearest neighbors.

Kumar, in [Kumar, 2020], introduces a novel mutliclass SVM approach to detect four types of subjects. i.e., people affected by hypothyroidism, hyperthyroidism, euthyroidism sick and euthyroidism healthy, on the usual UCI dataset.

Shen et al. [2016] propose a novel scheme to optimize the parameters of SVM by means of the fly optimization algorithm. The novel scheme is compared with other optimization algorithms and its efficiency is tested on four datasets, including the UCI one about thyroid.

Finally, in [Li et al., 2019b] a novel optimization technique for SVM applied to the UCI thyroid dataset is proposed. It is based on the teaching-learning algorithm and differential evolution and it permitted to SVM to reach better performance in comparison with other solutions.

Malignant nodules detection is the focus of the work in [Raghavendra et al., 2018]. The paper presents a computer-aided diagnosis system to detect thyroid malignant nodules by means of higher order spectral entropy features and using particle swarm optimization (PSO) and support vector machine (SVM) frameworks.

In [Prochazka et al., 2019], the authors propose a computer aided diagnosis system using direction independent features of ultrasound images of the thyroid gland to detect malignant nodules by means of Random Forest and SVM. Similarly, the authors of the study in [Raghavendra et al., 2017] try to propose

Deleted: s

an SVM-based computer aided diagnosis system, by exploiting fusion of Spatial Gray Level Dependence Features (SGLDF) and fractal textures to detect benign and malignant thyroid lesions in ultrasound images.

Deleted: with

In [Kaur et al., 2019], the authors proposed and IoT-based framework using SVM, Random Forest, Decision Trees, k-nearest neighbour, and artificial neural networks to be tested on different healthcare datasets, including on general thyroid disease.

Deleted: one about

Finally, in [Shankar et al., 2018] the authors propose again an optimal feature selection method, namely Improved Gray Wolf Optimization, to reduce the number of features for a multi-kernel SVM classifier applied to thyroid disease.

Deleted: i

Deleted: g

Deleted: w

Deleted: o

Shen et al. [2021] apply SVM to platelet RNA-seq data in order to differentiate different types of thyroid cancer and achieve an accuracy of 97%.

Khudhair Abbas [2021] tries to segment and classify ultrasound images of the thyroid gland using artificial neural networks and SVM classifiers. SVM demonstrated to achieve better performance than artificial neural network in detecting benign and malignant nodes with a 96.66% of sensitivity.

5.1.3 Decision Tree approaches

In Hayashi [2017], the author tries to create a white-box model to make prediction on the UCI thyroid datasets by exploiting the synergical effects of Recursive-Rule eXtraction (Re-RX) with J48graft in terms of rule definition in the form IF-THEN for predicting both hypo- and hyperthyroidism.

In Sidiq and Mutahar Aaqib [2019], the authors employ decision trees, random forest, vote ensemble, and stack ensemble in order to detect both hypo- and hyperthyroidism.

Hao et al. [2018] introduce a decision tree improved by MS-Apriori for the prognosis of lymph node metastasis (LNM) in patients with thyroid cancer. MS-Apriori is used to generate association rules considering rare items by multiple supports and fuzzy logic is introduced to map attribute values to different subintervals. The used dataset is made of Clinical-pathological data, obtained from the First Hospital of Jilin University.

Deleted: The authors of the contribution in

Jha et al. [2018] present a hybrid algorithm for healthcare data mining by using the Decision stump (DS), StackingC (SC), and voting methods to tackle the hyperthyroidism issue.

Deleted: The contribution in

Deleted: s

In Imbus et al. [2017], a Random Tree and a rule-based classifier (JRip) are used to detect primary hyperparathyroidism; while in Pan et al. [2016], random forest, together with principal component analysis and rotation transformation is used to detect a general thyroid disease by exploiting the UCI thyroid dataset.

Finally, in Zarin Mousavi et al. [2020] the authors apply computational methods based on decision trees, like ID3 and CHAID, SVM and multi-layer perceptrons, enriched with bagging and boosting techniques, to the identification of congenital hypothyroidism.

5.1.4 Rule-based approaches

In Bentaiba-Lagrid et al. [2020], a new amplification technique based on randomization for a system incorporating a structured case-base that speeds up case retrieval while supporting case retention is presented. The case base is segmented in a novel manner with new similarity functions based on features' weights in order to accelerate the retrieval of the case base reasoning. The system was applied to the detection of both hypo- and hyperthyroidism and compared with different machine learning methods.

In Asaad Sajadi et al. [2019], hypothyroidism is detected by means of a fuzzy rule-based expert system, which is proved to perform better than a logistic regression model on a real-world? dataset.

Finally, the authors of Kumari and Sharma [2019] draft a fuzzy logic-based expert system to tackle both hypo- and hyperthyroidism.

5.1.5 Neural Networks

Mahurkar and Gaikwad [2017] exploit artificial neural networks in conjunction with K-Means to normalize raw data for hypo- and hyperthyroidism.

In [Vivar et al., 2020], a guiding computer-aided diagnosis system, using a neural network with a dropout at the input layer, and integrated gradients of the trained network at test-time to attribute feature importance dynamically, is proposed. The technique is applied also to the UCI thyroid dataset to detect both hypo- and hyperthyroidism.

Similarly, in [Santos et al., 2019], the authors introduce a decision support system, to make general thyroid dysfunction assessment, based on an Artificial Neural Network and complemented by a novel approach to Knowledge Representation and Argumentation.

In [Yue et al., 2020], Fourier transform infrared spectroscopy was combined with three neural network models, namely multilayer perceptron, long-short-term memory network, and a convolutional neural network in order to detect hypo- and hyperthyroidism.

Chai [2020] tackles thyroid disease in general by means of knowledge graphs extracting the relationships between bio-medical entities for feeding a bidirectional long short-term memory network. This combination proved to have better diagnostic effects than other techniques on a image dataset from the university of Shanghai.

In [Zhang et al., 2020], the authors propose a synthetic data augmentation method based on progressive generative adversarial network in order to improve the performance in deep learning detection of hypo- and hyperthyroidism.

Yin et al. [2019] proposed a hybrid cutting network featuring a regional attribute cutting method for feature extraction and classification applied to a dataset of thyroid ultrasound images. The objective was to detect malignant thyroid nodules that could cause cancer.

In [Yi et al., 2017], the authors propose a novel diagnostic system to detect thyroid cancer, with a particular focus on thyroid nodule risk assessment. The method employs convolutional neural networks to analyze ultrasound images.

The contribution in [Lyu and Haque, 2018] embeds high dimensional RNA-Sequence data into bi-dimensional images and uses a convolutional neural network to detect various types of cancer, including thyroid cancer.

Deleted: among which also the one of the thyroid.

In [Li et al., 2019a], the authors employ deep convolutional neural networks to enhance the diagnostic accuracy of thyroid cancer through the analysis of sonographic images coming from clinical ultrasounds. Ultrasound images are also used in [Poudel et al., 2019] to feed convolutional neural networks for texture classification of anatomical structures of the thyroid for detecting general changes of its shape. Conversely, in [Moran et al., 2018] thermograms are exploited by convolutional neural networks for the early identification of thyroid nodules that can possibly cause cancers.

[Guo and Du, 2019] also use ultrasound images of the standard plane of the thyroid to evaluate its general status by means of deep convolutional neural networks. In particular, a 18-layer ResNet achieves the best results according to the authors.

Deleted: The contribution in

Deleted: exploits again

Deleted: is the one achieving

In [Shi et al., 2020], the authors integrate domain knowledge, extracted from standardized terminology, and deep learning (Auxiliary Classifier Generative Adversarial Network) into synthetic medical image augmentation to classify ultrasonography thyroid nodules.

In [Lu et al., 2020], hyperthyroidism is tackled, with a particular focus on its progression, by means of enhanced LSTMs with an adaptive loss function. The analyzed data regard blood test information in the early stage from a Shangai hospital.

Cordes et al. [2021] detect thyroid malignant nodules using artificial neural networks on ultrasonographic characteristics obtaining an accuracy of 84.4%. Similarly, Jin et al. [2021] analyze clinical ultrasound imaging data in five hospitals in China through artificial neural networks obtaining performance between 80% and 90%. Ahmed et al. [2022] use artificial deep neural networks on the concatenation of 6 databases containing data collected from Garvan Institute in Sydney, Australia to discover thyroid cancer. The main and only shown result is the accuracy of 98%; and, the results and the optimization process are not described in detail.

Deleted: try to

Deleted: try to exploit

Deleted: indeed

Ananthi et al. [2022] and Khan [2021] exploit convolutional neural networks to try to detect the onset of thyroid dysfunctions. The former employs x-ray images to prevent hypo- and hyperthyroidism as well as malignant nodules and other dysfunctions reaching a 99% of accurate prediction, while the latter only focus on hypothyroidism and achieve 98% accuracy.

Deleted: of

Liu et al. [2021b] tackle the issue of malignant nodes identification using convolutional neural networks as well. In this case, they apply information fusion techniques on ultrasound images and radio frequency signals, achieving better results than using only ultrasound images in detecting malignant thyroid nodules.

In [Santillan et al., 2021] and [Song et al., 2022] the focus is on the detection of malignant thyroid nodules which could possibly generate cancer. The former apply convolutional and recurrent neural networks to Fourier Transform infrared spectroscopy data, achieving accuracy of 98.06%, while the latter employ a feature-enhanced dual branch convolutional neural network on

Deleted: always

Deleted: a top

Deleted: value equal to

ultrasound images of the thyroid gland obtaining the best mean average precision of identification equal to 92.5%.

Chu et al. [2021] face the issue of malignant thyroid nodule detection too, employing a mark-guided ultrasound deep network segmentation model which, in turn, is based on different types of convolutional neural networks. They achieve a segmentation accuracy equal to 97.85%, improving the outcomes of other standard convolutional neural networks.

Yang et al. [2021] use convolutional neural networks on thyroid scintigrams to detect general dysfunctions of the gland and report an accuracy of 92.73%. Similarly, Pi et al. [2022] exploit convolutional neural networks and the fusion of deep and handcrafted features from thyroid scintigraphy to early detect thyroid dysfunctions. They were able to reach an accuracy equal to 91.18% and an f-measure equal to 88.11%.

Deleted: exploit

Deleted: reaching an

Deleted: equal to

In [Zhao et al., 2022], semantic consistency generative adversarial network are employed to detect malignant thyroid nodules using ultrasound data. The proposed method is claimed to reach an accuracy equal to 94.30% and an Area Under the Curve equal to 97.02%.

Hosseinzadeh et al. [2021] use a multiple multilayer perceptron neural network to identify hypo- and hyperthyroidism in the context of Internet of Medical Things, reaching an accuracy of 99%. Saktheeswari and Balasubramanian [2021] exploit an autoencoder-based neural network using also a Multi-layer Tree-based State Machine to detect malignant thyroid nodules as well as hypo- and hyperthyroidism with a final mean accuracy equal to 98.90%.

Finally, Ai et al. [2022] apply a recent type of neural networks, i.e., capsule networks, on ultrasonic thyroid images to detect possible thyroid cancer traces achieving a top accuracy equal to 81.06%.

5.1.6 Ensembles

In [Yadav and Pal, 2022], an ensemble of different machine learning techniques is employed to detect thyroid hormone disease. The ensemble uses Boosting, Bagging, Stacking, and Voting and it is aimed at identifying hypothyroidism patients.

In [Alam et al., 2020], the authors present a novel dynamic ensemble learning of neural networks. It provides an automatic design of the ensemble, a maintaining of accuracy and diversity of the composing neural networks, and very few parameters to be designed by the user. The technique is successfully employed to detect hypothyroidism, hyperthyroidism, and euthyroidism disease.

Deleted: made

Deleted: t

Priyadharsini and Sasikala [2022] exploit Adaboost and Bagging as ensemble machine learning methods to correctly detect hypo- and hyperthyroidism. Bagging resulted in better performance as regards all the main metrics: accuracy 99.20%, precision 99.9%, f-measure 99.8%.

Finally, Yadav and Pal [2022] exploit Boosting, Bagging, Stacking, and Voting ensembles to detect hypothyroidism achieving a top accuracy equal to 99.86% and a top recall equal to 99.88%.

5.1.7 Other Learning Approaches

In Devi and Anita [2018], a semi-supervised approach is used to identify a general thyroid disease by analyzing tongue images.

In Priya and Manavalan [2018], a Weighted Extreme Learning Machine technique hybridized with Invasive Weed optimization is used to detect general thyroid disease.

Conversely, in Pavya and Srinivasan [2017], filter-based and wrapper-based feature selection methods are applied to four classifiers, namely, MultiLayer Perceptron Back Propagation Neural Network, Support Vector Machine, and Extreme Learning Machine, in order to detect both hypo- and hyperthyroidism.

In Ma et al. [2018], the authors introduce a novel hybrid diagnosis system, integrating local fisher discriminant analysis and kernelized extreme learning machine method for thyroid disease diagnosis (hypo- and hyper thyroidism).

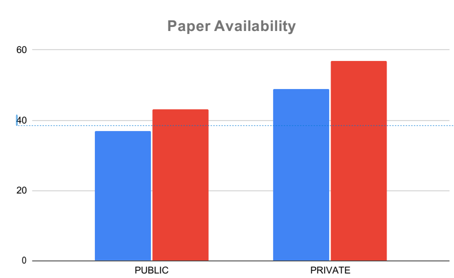
Finally, in Baccour [2018], a new classification system for both hypo- and hyperthyroidism and based on fused VIKOR and TOPSIS is proposed.

5.2 RQ2: What datasets about thyroid diseases are used in the considered AI solutions?

In this section, we describe the main characteristics of the datasets used in the surveyed papers.

In Figure 6, we show the availability of the datasets used in the papers we considered in this systematic review. As one can see, the number of private or non-available datasets is high (56.98%) compared with the number of publicly available ones.

Deleted: preponderant



Commented [A25]: y-axis title missing.

Figure 6. Availability of the datasets in the surveyed papers (blue: number of papers, red: percentage).

Moreover, the most used dataset is the UCI one[5] (25 times), followed by far by private datasets of Shangai hospitals (5 times) and then by the public KEEL dataset[6] (4 times).

Commented [A26]: Please put full form also.

Finally, in Table 4, we present some statistics about the main characteristics of the datasets used in the surveyed papers. They show that the average number of instances per dataset is about 5705 samples, while the average number of features is about 26,000, but in this case we have to consider that for image

[5]https://archive.ics.uci.edu/ml/datasets/Thyroid+Disease
[6]https://sci2s.ugr.es/keel/dataset.php?cod=66

Table 4. Statistics of the number of instances and of features in the considered datasets.

	Instances	Features
Average	5705	25674
Maximum	92062	711680
Minimum	92	4

datasets the number of features was considered equal to the number of pixels, thus making the statistics increase very much.

5.3 RQ3: What data types are used to detect and classify thyroid diseases using the considered AI techniques?

In this section, we analyze the datasets types considered in the surveyed papers.

In Figure 7, we show a pie chart to highlight the different percentages of data used in the surveyed papers. As it can be easily inferred, the majority of the considered features (51.22%) is made of clinical data, e.g., data from the blood analysis, while 28.05% of the surveyed papers use images of the thyroid, and only a small percentage (12.2%) exploit both types of data.

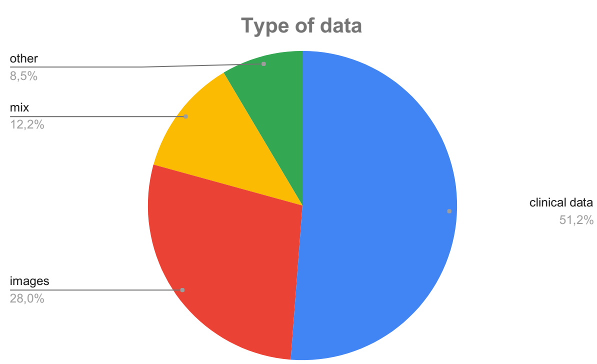


Figure 7. Type of data used in the considered datasets.

Table 5 shows the occurrences of the ten most used features in the surveyed articles. As it is evident, the most used feature is the thyroid-stimulating hormone (TSH) with 39 occurrences, followed by triiodothyronine (T3) hormone, the age, and the gender of the subject. Other very used features regard the pixel value (20), the only feature used whenever image datasets are employed, as well as the thyroxine, also called T4, like FTI (Free T4 Index) and T4U (T4 Uptake). Pregnancy status is the least important feature in the study context.

Deleted: more deeply

Commented [A27]: Ambiguous.

Deleted: Finally, the least positions in the top ten¶
522 list are occupied by the p

Formatted: Body Text

Deleted: .¶

Table 5. Occurrences of the top ten features used in the surveyed papers.

Feature	Occurrences
TSH	39
T3	28
Age	28
Gender	27
Pixel value	20
TT4	19
FTI	18
T4U	18
Pregnancy	14
T4	13

6 DISCUSSION

In this section, we perform a brief discussion about the main outcomes of the systematic review, and some important open research directions.

One of the first results is that a great majority of works focus on supervised methods. Therefore, the development of novel

and effective techniques employing also unsupervised or semi-supervised approaches is surely a need in the application of artificial intelligence techniques to the automated classification of thyroid dysfunctions and diseases.

As regards the tackled diseases, there is a clear prevalence of hypo- and hyperthyroidism. Therefore, more focused scientific work, especially regarding euthyroid pathology and cancers, should be carried out in

order to foster use of AI techniques in the detection of thyroid dysfunction.

The average accuracy in detecting thyroid-related diseases is far from optimal; thus, given also higher use of neural networks among the surveyed techniques, a greater focus on deep neural networks and the careful optimization of their parameters should be the topics of future research papers.

As regards the datasets, too few thereof are publicly available, and among these, the usage of the UCI dataset, by now rather obsolete (1987), is dominant. The need of novel, public, and updated datasets is, thus, compelling, together with the application of proper feature selection and feature reduction techniques.

Finally, in the surveyed datasets, the usage of clinical data is prevalent, thus, also in the view of a greater usage of various deep learning techniques, a greater development of public image datasets to be used with proper deep convolutional neural networks should be encouraged.

Commented [A28]: Please provide abbreviations in table footnotes.

Deleted: as well
525 as about

Formatted: Space Before: 3.25 pt

Deleted: the fact

Deleted: thus
527 techniques requiring the labeling in advance of the whole dataset…

Formatted: Space Before: 6.45 pt, Tab stops: 2.49 cm, Left

Commented [A29]: More justification is needed. Perhaps, the nature of problem necessitates the use of supervised learning?

Deleted: researches

Deleted: a more widespread usage

Deleted: prevention of such important dysfunctions
534 of the thyroid gland.

Formatted: Body Text, Indent: Left: 0 cm, Space Before: 6.4 pt

Deleted: still

Deleted: the preponderant usage

Commented [A30]: Please use simple sentence.

7 CONCLUSIONS

In this paper, we have performed a detailed systematic review about the application of artificial intelligence techniques to the study, analysis, and classification of various thyroid-related diseases and dysfunctions.

First of all, we have summarized some similar reviews and pointed out the necessity of the systematic review we have carried out.

Subsequently, we have detailed the research process we employed, describing the research questions, the used databases and the employed queries. We have also described in detail the different filtering phases that led us to find out the 74 analyzed papers.

In the end, we have presented and discussed in depth the results of the systematic review, in order to answer to the proposed research questions: the tackled thyroid diseases, the used artificial intelligence techniques, the used datasets, the types of the thyroid-related features.

A final discussion has pointed out the main open research directions, as well as the issues still to be solved in order to make artificial intelligence a viable solution for the quick diagnosis and classification of any thyroid-related disease or dysfunction.

Commented [A31]: This looks more like methodology summary, not a conclusion per se. Please write the findings of your systematic review in concise terms.

ACKNOWLEDGMENTS

The work of Riccardo Pecori has been partly supported by the PON R&I 2014-2020 "AIM: Attraction and International Mobility" project. The work of Mario Luca Bernardi has been supported by the Italian Ministry of Education and Research (MIUR), in the framework of the "Departments of Excellence" project. The supporting institutions had no role in the study design, in the collection, analysis and interpretation of data, in the writing of the report, and in the decision to submit the article for publication.

AUTHORS' CONTRIBUTIONS

Lerina Aversano: Conceptualization, Methodology, Supervision. Mario Luca Bernardi: Validation, Writing - Review & Editing. Marta Cimitile and Riccardo Pecori: Conceptualization, Methodology, Supervision, Writing - Review & Editing. Riccardo Pecori and Andrea Maiellaro: Formal analysis, Investigation, Resources, Data curation, Writing - Original Draft.

REFERENCES

Abdolali, F., Shahroudnejad, A., Hareendranathan, A. R., Jaremko, J. L., Noga, M., and Punithakumar, K. (2020). A systematic review on the role of artificial intelligence in sonographic diagnosis of thyroid cancer: Past, present and future. *CoRR*, abs/2006.05861.

Ahmed, I., Mohiuddin, R., Muqeet, M. A., Kumar, J. A., and Thaniserikaran, A. (2022). Thyroid cancer detection using deep neural network. In *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, pages 166–169.

Ahmed, J. and Soomrani, M. A. R. (2016). Tdtd: Thyroid disease type diagnostics. In *2016 International Conference on Intelligent Systems Engineering (ICISE)*, pages 44–50.

Ai, X., Zhuang, J., Wang, Y., Wan, P., and Fu, Y. (2022). Rescaps: an improved capsule network and its application in ultrasonic image classification of thyroid papillary carcinoma. *Complex & Intelligent Systems*, 8(3):1865–1873.

Alam, K., Siddique, N., and Adeli, H. (2020). A dynamic ensemble learning algorithm for neural networks. *Neural Computing and Applications*, 32.

Ananthi, N., Kumaran, K., Duraiarumugam, S., Lavanya, P., Ganesan, S., Sathiya, G. R., and Surya, J. (2022). Detecting six different types of thyroid diseases using deep learning approaches. In *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, pages 1–8.

Asaad Sajadi, N., Borzouei, S., Mahjub, H., and Farhadian, M. (2019). Diagnosis of hypothyroidism using a fuzzy rule-based expert system. *Clinical Epidemiology and Global Health*, 7(4):519–524.

Aversano, L., Bernardi, M. L., Cimitile, M., and Pecori, R. (2021). Deep neural networks ensemble to detect covid-19 from ct scans. *Pattern Recognition*, 120:108135.

Baccour, L. (2018). Amended fused topsis-vikor for classification (atovic) applied to some uci data sets. *Expert Systems with Applications*, 99:115–125.

Bentaiba-Lagrid, M. B., Bouzar-Benlabiod, L., Rubin, S. H., Bouabana-Tebibel, T., and Hanini, M. R. (2020). A case-based reasoning system for supervised classification problems in the medical field. *Expert Systems with Applications*, 150:113335.

Bini, F., Pica, A., Azzimonti, L., Giusti, A., Ruinelli, L., Marinozzi, F., and Trimboli, P. (2021). Artificial intelligence in thyroid field—a comprehensive review. *Cancers*, 13(19).

Chai, X. (2020). Diagnosis method of thyroid disease combining knowledge graph and deep learning. *IEEE Access*, 8:149787–149795.

Chandel, K., Kunwar, V., Sabitha, S., Choudhury, T., and Mukherjee, S. (2016). A comparative study on thyroid disease detection using k-nearest neighbor and naive bayes classification techniques. *CSI Transactions on ICT*, 4:313–319.

Chandio, J. A., Sahito, A., Soomrani, M. A. R., and Abbasi, S. A. (2016). Tdv: Intelligent system for thyroid disease visualization. In *2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, pages 106–112.

Chaubey, G., Bisen, D., Arjaria, S., and Yadav, V. (2021). Thyroid disease prediction using machine learning approaches. *National Academy Science Letters*, 44(3):233–238.

Chen, J., You, H., and Li, K. (2020). A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images. *Computer Methods and Programs in Biomedicine*, 185:105329.

Chu, C., Zheng, J., and Zhou, Y. (2021). Ultrasonic thyroid nodule detection method based on u-net network. *Computer Methods and Programs in Biomedicine*, 199:105906.

Cordes, M., Götz, T. I., Lang, E. W., Coerper, S., Kuwert, T., and Schmidkonz, C. (2021). Advanced thyroid carcinomas: neural network analysis of ultrasonographic characteristics. *Thyroid Research*, 14(1):16.

Devi, G. and Anita, E. (2018). A novel semi supervised learning algorithm for thyroid and ulcer classification in tongue image. *Cluster Computing*, pages 1–13.

Duggal, P. and Shukla, S. (2020). Prediction of thyroid disorders using advanced machine learning techniques. In *2020 10th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 670–675.

Garg, A. and Mago, V. (2021). Role of machine learning in medical research: A survey. *Computer Science Review*, 40:100370.

Guo, M. and Du, Y. (2019). Classification of thyroid ultrasound standard plane images using resnet-18 networks. In *2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID)*, pages 324–328.

Hao, Y., Zuo, W., Shi, Z., Yue, L., Xue, S., and He, F. (2018). Prognosis of thyroid disease using ms-apriori improved decision tree. In Liu, W., Giunchiglia, F., and Yang, B., editors, *Knowledge Science, Engineering and Management*, pages 452–460, Cham. Springer International Publishing.

Hasanzad, M., Aghaei Meybodi, H. R., Sarhangi, N., and Larijani, B. (2022). Artificial intelligence perspective in the future of endocrine diseases. *Journal of Diabetes & Metabolic Disorders*, 21(1):971–978.

Hayashi, Y. (2017). Synergy effects between grafting and subdivision in re-rx with j48graft for the diagnosis of thyroid disease. *Knowledge-Based Systems*, 131:170–182.

Hosseinzadeh, M., Ahmed, O. H., Ghafour, M. Y., Safara, F., hama, H. k., Ali, S., Vo, B., and Chiang, H.-S. (2021). A multiple multilayer perceptron neural network with an adaptive learning algorithm for thyroid disease diagnosis in the internet of medical things. *The Journal of Supercomputing*, 77(4):3616–3637.

Houari, R., Bounceur, A., Kechadi, M.-T., Tari, A.-K., and Euler, R. (2016). Dimensionality reduction in data mining: A copula approach. *Expert Systems with Applications*, 64:247–260.

Imbus, J. R., Randle, R. W., Pitt, S. C., Sippel, R. S., and Schneider, D. F. (2017). Machine learning to identify multigland disease in primary hyperparathyroidism. *Journal of Surgical Research*, 219:173–179.

Jha, S. K., Paramasivam, P., Pan, Z., and Wang, J. (2018). Decision stump and stackingc-based hybrid algorithm for healthcare data classification. In Sun, X., Pan, Z., and Bertino, E., editors, *Cloud Computing and Security*, pages 205–216, Cham. Springer International Publishing.

Jin, Z., Zhu, Y., Zhang, S., Xie, F., Zhang, M., Guo, Y., Wang, H., Zhu, Q., Cao, J., and Luo, Y. (2021). Diagnosis of thyroid cancer using a ti-rads-based computer-aided diagnosis system: a multicenter retrospective study. *Clinical Imaging*, 80:43–49.

Juneja, K. (2022). Expanded and filtered features based elm model for thyroid disease classification. *Wireless Personal Communications*, 126(2):1805–1842.

Kaur, P., Kumar, R., and Kumar, M. (2019). A healthcare monitoring system using random forest and internet of things (iot). *Multimedia Tools Appl.*, 78(14):19905–19916.

Khan, T. (2021). Application of two-class neural network-based classification model to predict the onset of thyroid disease. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pages 114–118.

Khudhair Abbas, S. (2021). Thyroid tissue segmentation and classification in ultrasound image of artificial intelligence. *Materials Today: Proceedings*.

Kishor, A. and Chakraborty, C. (2021). Artificial intelligence and internet of things based healthcare 4.0 monitoring system. *Wireless Personal Communications*.

Kitchenham, B. (2004). Procedures for Performing Systematic Reviews. *Keele, UK, Keele Univ.*, pages 1–33.

Kumar, H. H. S. (2020). A novel approach of svm based classification on thyroid disease stage detection. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pages 836–841.

Kumar, Y., Gupta, S., Singla, R., and Hu, Y.-C. (2022a). A systematic review of artificial intelligence techniques in cancer prediction and diagnosis. *Archives of Computational Methods in Engineering*, 29(4):2043–2070.

Kumar, Y., Koul, A., Singla, R., and Ijaz, M. F. (2022b). Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *Journal of Ambient Intelligence and Humanized Computing*.

Kumari, P. and Sharma, S. (2019). Fuzzy based medicine recommendation system: An example of thyroid medicine. ICAICR '19, New York, NY, USA. Association for Computing Machinery.

Kwak, G. H.-J. and Hui, P. (2019). DeepHealth: Deep Learning for Health Informatics. *ArXiv*, abs/1909.00384.

Li, X., Zhang, S., Zhang, Q., Wei, X., Pan, Y., Zhao, J., Xin, X., Qin, C., Wang, X., Li, J., Yang, F., Zhao, Y., Yang, M., Wang, Q., Zheng, Z., Zheng, X., Yang, X., Whitlow, C. T., Gurcan, M. N., Zhang, L., Wang, X., Pasche, B. C., Gao, M., Zhang, W., and Chen, K. (2019a). 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.

Li, Z., Qin, J., Zhang, X., and Wan, Y. (2019b). A hybrid intelligent framework for thyroid diagnosis. In Ning, H., editor, *Cyberspace Data and Intelligence, and Cyber-Living, Syndrome, and Health*, pages 441–451, Singapore. Springer Singapore.

Liu, P.-r., Lu, L., Zhang, J.-y., Huo, T.-t., Liu, S.-x., and Ye, Z.-w. (2021a). Application of artificial intelligence in medicine: An overview. *Current Medical Science*, 41(6):1105–1115.

Liu, Z., Zhong, S., Liu, Q., Xie, C., Dai, Y., Peng, C., Chen, X., and Zou, R. (2021b). Thyroid nodule recognition using a joint convolutional neural network with information fusion of ultrasound images and radiofrequency data. *European Radiology*, 31(7):5001–5011.

Longbottom, J. and Macnab, R. (2014). Thyroid disease and thyroid surgery. *Anaesthesia & Intensive Care Medicine*, 15:458–464.

Lu, H., Wang, M., Zhao, W., Su, T., and Yang, J. (2020). Hyperthyroidism progress prediction with enhanced lstm. In Wang, X., Zhang, R., Lee, Y.-K., Sun, L., and Moon, Y.-S., editors, *Web and Big Data*, pages 209–217, Cham. Springer International Publishing.

Lyu, B. and Haque, A. (2018). Deep learning based tumor type classification using gene expression data. In *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, BCB '18, page 89–96, New York, NY, USA. Association for Computing Machinery.

Ma, C., Guan, J., Zhao, W., and Wang, C. (2018). An efficient diagnosis system for thyroid disease based on enhanced kernelized extreme learning machine approach. In Xiao, J., Mao, Z.-H., Suzumura, T., and Zhang, L.-J., editors, *Cognitive Computing – ICCC 2018*, pages 86–101, Cham. Springer International Publishing.

Ma, L., Ma, C., Liu, Y., and Wang, X. (2019). Thyroid diagnosis from spect images using convolutional neural network with optimization. *Computational Intelligence and Neuroscience*, 2019:6212759.

Mahurkar, K. K. and Gaikwad, D. (2017). Normalization using improvised k-means applied in diagnosing thyroid disease with ann. In *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, pages 579–583.

Mendoza, A. M. and Hernandez, R. M. (2021). Application of data mining techniques in diagnosing various thyroid ailments: A review. In *2021 13th International Conference on Information & Communication Technology and System (ICTS)*, pages 207–212.

Monaco, F. (2003). Classification of Thyroid Diseases: Suggestions for a Revision. *The Journal of Clinical Endocrinology & Metabolism*, 88(4):1428–1432.

Moran, M. B. H., Conci, A., González, J. R., Araújo, A. S., Fiirst, W. G., Damião, C. P., Lima, G. A. B., and Filho, R. A. d. C. (2018). Identification of thyroid nodules in infrared images by convolutional neural networks. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7.

Pan, Q., Zhang, Y., Zuo, M., Xiang, L., and Chen, D. (2016). Improved ensemble classification method of thyroid disease based on random forest. In *2016 8th International Conference on Information Technology in Medicine and Education (ITME)*, pages 567–571.

Parmar, B. S. and Mehta, M. A. (2020). Computer-aided diagnosis of thyroid dysfunction: A survey. In Bellatreche, L., Goyal, V., Fujita, H., Mondal, A., and Reddy, P. K., editors, *Big Data Analytics*, pages 164–189, Cham. Springer International Publishing.

Pasha, S. J. and Mohamed, E. (2020). Ensemble gain ratio feature selection (egfs) model with machine learning and data mining algorithms for disease risk prediction. In *2020 International Conference on Inventive Computation Technologies (ICICT)*, pages 590–596.

Pavya, K. and Srinivasan, B. (2017). Feature selection algorithms to improve thyroid disease diagnosis. In *2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT)*, pages 1–5.

Peya, Z. J., Chumki, M. K. N., and Zaman, K. M. (2021). Predictive analysis for thyroid diseases diagnosis using machine learning. In *2021 International Conference on Science & Contemporary Technologies (ICSCT)*, pages 1–6.

Pi, Y., Yang, P., Wei, J., Zhao, Z., Cai, H., and Yi, Z. (2022). Fusing deep and handcrafted features for intelligent recognition of uptake patterns on thyroid scintigraphy. *Knowledge-Based Systems*, 236:107531.

Poudel, P., Illanes, A., Sadeghi, M., and Friebe, M. (2019). Patch based texture classification of thyroid ultrasound images using convolutional neural network. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5828–5831.

Priya, S. and Manavalan, R. (2018). Tuning the parameters of weighted elm using iwo and bat algorithm to improve the classification performance. In *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on*, pages 547–552.

Priyadharsini, D. and Sasikala, S. (2022). Efficient thyroid disease prediction using features selection and meta-classifiers. In *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, pages 1236–1243.

Prochazka, A., Gulati, S., Holinka, S., and Smutek, D. (2019). Patch-based classification of thyroid nodules in ultrasound images using direction independent features extracted by two-threshold binary decomposition. *Computerized Medical Imaging and Graphics*, 71:9–18.

Qin, H., Que, Q., Lin, P., Li, X., Wang, X.-r., He, Y., Chen, J.-q., and Yang, H. (2021). Magnetic resonance imaging (mri) radiomics of papillary thyroid cancer (ptc): a comparison of predictive performance of multiple classifiers modeling to identify cervical lymph node metastases before surgery. *La radiologia medica*, 126(10):1312–1327.

Raghavendra, U., Gudigar, A., Maithri, M., Gertych, A., Meiburger, K. M., Yeong, C. H., Madla, C., Kongmebhol, P., Molinari, F., Ng, K. H., and Acharya, U. R. (2018). Optimized multi-level elongated quinary patterns for the assessment of thyroid nodules in ultrasound images. *Computers in Biology and Medicine*, 95:55–62.

Raghavendra, U., Rajendra Acharya, U., Gudigar, A., Hong Tan, J., Fujita, H., Hagiwara, Y., Molinari, F., Kongmebhol, P., and Hoong Ng, K. (2017). Fusion of spatial gray level dependency and fractal texture features for the characterization of thyroid lesions. *Ultrasonics*, 77:110–120.

Raisinghani, S., Shamdasani, R., Motwani, M., Bahreja, A., and Raghavan Nair Lalitha, P. (2019). Thyroid prediction using machine learning techniques. In Singh, M., Gupta, P., Tyagi, V., Flusser, J., Ören, T., and Kashyap, R., editors, *Advances in Computing and Data Sciences*, pages 140–150, Singapore. Springer Singapore.

Rao, A. and Renuka, B. (2020). A machine learning approach to predict thyroid disease at early stages of diagnosis. In *2020 IEEE International Conference for Innovation in Technology (INOCON)*, pages 1–4.

Razia, S., Siva Kumar, P., and Rao, A. S. (2020). *Machine Learning Techniques for Thyroid Disease Diagnosis: A Systematic Review*, pages 203–212. Springer International Publishing, Cham.

Riajuliislam, M., Rahim, K. Z., and Mahmud, A. (2021). Prediction of thyroid disease(hypothyroid) in early stage using feature selection and classification techniques. In *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, pages 60–64.

Saktheeswari, M. and Balasubramanian, T. (2021). Multi-layer tree liquid state machine recurrent auto encoder for thyroid detection. *Multimedia Tools and Applications*, 80(12):17773–17783.

Santillan, A., Tomas, R. C., Bangaoil, R., Lopez, R., Gomez, M. H., Fellizar, A., Lim, A., Abanilla, L., Ramos, M. C., Guevarra, L., and Albano, P. M. (2021). Discrimination of malignant from benign thyroid lesions through neural networks using ftir signals obtained from tissues. *Analytical and Bioanalytical Chemistry*, 413(8):2163–2180.

Santos, S., Martins, M. R., Vicente, H., Barroca, M. G., Calisto, F., Gama, C., Ribeiro, J., Machado, J., Ávidos, L., Araújo, N., Dias, A., and Neves, J. (2019). A many-valued empirical machine for thyroid dysfunction assessment. In Cortez, P., Magalhães, L., Branco, P., Portela, C. F., and Adão, T., editors, *Intelligent Technologies for Interactive Entertainment*, pages 47–57, Cham. Springer International Publishing.

Shahid, A. H., Singh, M. P., Raj, R. K., Suman, R., Jawaid, D., and Alam, M. (2019). A study on label tsh, t3, t4u, tt4, fti in hyperthyroidism and hypothyroidism using machine learning techniques. In *2019 International Conference on Communication and Electronics Systems (ICCES)*, pages 930–933.

Shankar, K., Lakshmanaprabu, S., Gupta, D., Maseleno, A., and Albuquerque, V. (2018). Optimal feature-based multi-kernel svm approach for thyroid disease classification. *The Journal of Supercomputing*, 76:1128–1143.

Sharifi, Y., Bakhshali, M. A., Dehghani, T., DanaiAshgzari, M., Sargolzaei, M., and Eslami, S. (2021). Deep learning on ultrasound images of thyroid nodules. *Biocybernetics and Biomedical Engineering*, 41(2):636–655.

Shen, L., Chen, H., Yu, Z., Kang, W., Zhang, B., Li, H., Yang, B., and Liu, D. (2016). Evolving support vector machines using fruit fly optimization for medical data classification. *Knowledge-Based Systems*, 96:61–75.

Shen, Y., Lai, Y., Xu, D., Xu, L., Song, L., Zhou, J., Song, C., and Wang, J. (2021). Diagnosis of thyroid neoplasm using support vector machine algorithms based on platelet rna-seq. *Endocrine*, 72(3):758–783.

Shi, G., Wang, J., Qiang, Y., Yang, X., Zhao, J., Hao, R., Yang, W., Du, Q., and Kazihise, N. G.-F. (2020). Knowledge-guided synthetic medical image adversarial augmentation for ultrasonography thyroid nodule classification. *Computer Methods and Programs in Biomedicine*, 196:105611.

Sidiq, U. and Mutahar Aaqib, S. (2019). Disease diagnosis through data mining techniques. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pages 275–280.

Song, R., Zhu, C., Zhang, L., Zhang, T., Luo, Y., Liu, J., and Yang, J. (2022). Dual-branch network via pseudo-label training for thyroid nodule detection in ultrasound image. *Applied Intelligence*, 52(10):11738–11754.

Tyagi, A., Mehra, R., and Saxena, A. (2018). Interactive thyroid disease prediction system using machine learning technique. In *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pages 689–693.

Vivar, G., Mullakaeva, K., Zwergal, A., Navab, N., and Ahmadi, S.-A. (2020). Peri-diagnostic decision support through cost-efficient feature acquisition at test-time. In Martel, A. L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M. A., Zhou, S. K., Racoceanu, D., and Joskowicz, L., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pages 572–581, Cham. Springer International Publishing.

Wilson, B. S., Tucci, D. L., Moses, D. A., Chang, E. F., Young, N. M., Zeng, F.-G., Lesica, N. A., Bur, A. M., Kavookjian, H., Mussatto, C., Penn, J., Goodwin, S., Kraft, S., Wang, G., Cohen, J. M., Ginsburg, G. S., Dawson, G., and Francis, H. W. (2022). Harnessing the power of artificial intelligence in otolaryngology and the communication sciences. *Journal of the Association for Research in Otolaryngology*, 23(3):319–349.

Witten, Ian. H., F. E. H. M. A. (2011). *Data Mining, Practical Machine Learning Tools*. Elsevier.

Yadav, D. C. and Pal, S. (2022). Thyroid prediction using ensemble data mining techniques. *International Journal of Information Technology*, 14(3):1273–1283.

Yang, P., Pi, Y., He, T., Sun, J., Wei, J., Xiang, Y., Jiang, L., Li, L., Yi, Z., Zhao, Z., and Cai, H. (2021). Automatic differentiation of thyroid scintigram by deep convolutional neural network: a dual center study. *BMC Medical Imaging*, 21(1):179.

Yi, Z., Hao, A., Song, W., Li, H., and Li, B. (2017). A novel intelligent thyroid nodule diagnosis system over ultrasound images based on deep learning. In *2017 International Conference on Virtual Reality and Visualization (ICVRV)*, pages 150–155.

Yin, J., Song, R., Wang, J., Zhu, C., Yang, S., and Yang, J. (2019). Hybrid regional feature cutting network for thyroid ultrasound images classification. In *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*, pages 1792–1797.

Yue, F., Chen, C., Yan, Z., Chen, C., Guo, Z., Zhang, Z., Chen, Z., Zhang, F., and Lv, X. (2020). Fourier transform infrared spectroscopy combined with deep learning and data enhancement for quick diagnosis of abnormal thyroid function. *Photodiagnosis and Photodynamic Therapy*, 32:101923.

Zarin Mousavi, S. S., Mohammadi Zanjireh, M., and Oghbaie, M. (2020). Applying computational classification methods to diagnose congenital hypothyroidism: A comparative study. *Informatics in Medicine Unlocked*, 18:100281.

Zhang, H., Huang, Z., and Lv, Z. (2020). Medical image synthetic data augmentation using gan. In *Proceedings of the 4th International Conference on Computer Science and Application Engineering*, CSAE 2020, New York, NY, USA. Association for Computing Machinery.

Zhao, J., Zhou, X., Shi, G., Xiao, N., Song, K., Zhao, J., Hao, R., and Li, K. (2022). Semantic consistency generative adversarial network for cross-modality domain adaptation in ultrasound thyroid nodule classification. *Applied Intelligence*, 52(9):10369–10383.