A systematic review on artificial intelligence techniques for detecting 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] Dept. of Law and digital Society, UnitelmaSapienza University, Rome, Italy

[3] Institute of Materials for Electronics and Magnetism, National Research Council, Parma, Italy

[4] SMARTEST Research Centre, eCampus University, Novedrate (CO), Italia

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 an example of illnesses that can be effectively faced if discovered quite early. Detecting thyroid diseases is crucial in order to treat patients effectively and promptly, by saving lives and reducing healthcare costs. This work aims at systematically reviewing and analyzing the literature on 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 and taxonomies in order to highlight pros and cons of the most recent research in the field. After a careful selection process, we selected and reviewed 72 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 The average accuracy of detection is high (96.84%), but the usage of private and outdated datasets with a majority of clinical data is very common. 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.

A Systematic Review on Artificial Intelligence Techniques for Detecting 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]**Dept. of Law and Digital Society, UnitelmaSapienza University, P.za Sassari 4, Rome, 00161, Italy**

[3]**SMARTEST Research Centre, 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 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 an example of illnesses that can be effectively faced if discovered quite early. Detecting thyroid diseases is crucial in order to treat patients effectively and promptly, by saving lives and reducing healthcare costs.

This work aims at systematically reviewing and analyzing the literature on 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 and taxonomies in order to highlight pros and cons of the most recent research in the field.

After a careful selection process, we selected and reviewed 72 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 (96.84%), but the usage of private and outdated datasets with a majority of clinical data is very common.

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.

1 INTRODUCTION

Thyroid is an important human endocrine organ, positioned in the anterior part of the neck and secretes the hormone which regulates the human metabolism. A thyroid disorder takes place whenever 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. With respect to several other diseases, thyroid diseases affect about 200 million people worldwide. Moreover, an estimated 40 percent of the world population is at risk of iodine nutrient deficiency, essential for the production of thyroid hormones, that is the major cause of morbidity worldwide [Keestra et al., 2020; Aversano et al., 2021a]. For this reason, it is very important to diagnose thyroid diseases at their early stages and take precautions to avoid the most dangerous conditions. There are several approaches to diagnose thyroid disorder such as clinical test evaluation, imaging inspection, blood analysis, and tissue biopsy. These approaches require the work of doctors and are not foolproof because the process of diagnosis of thyroid disorder from the laboratory analysis is quite complex and requires the doctors' extensive knowledge and experience. For this reason, in the last years, several research studies, using Artificial Intelligence (AI) techniques to detect specifically various thyroid diseases [Ma et al., 2019; Kwak and Hui, 2019], have been carried out. Indeed, AI has been extensively used to solve problems in the healthcare field [Kwak and Hui, 2019; Aversano et al., 2021b] and has often 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. In this direction, the proposed systematic review aims to provide some useful insights for all the clinicians and researchers focusing on the topic of thyroid disease detection by identifying the main adopted AI techniques, the most adopted datasets and the types of data useful to perform thyroid disease detection.

The target 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 summarized in the following:

- summarizing the most recent AI solutions linked to the early detection of thyroid diseases (Research Question RQ1);

- identifying the used datasets to apply AI solutions for the early detection 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

68 and classify the considered articles, in a way suitable to answer the research questions. Finally, Section 6

69 reports a discussion of the obtained findings, also highlighting the current research gaps, while Section 7

70 concludes the article with an overview of the systematic review and of the obtained results, drawing also

71 some possible future developments in the field.

2 BACKGROUND

73 The primary function the thyroid gland is the production of the triiodothyronine (T3) and thyroxine (T4)

74 hormones. These hormones travel through the body and help the regulation of the metabolism, while also

75 aiding brain development, digestive function, muscle control, and mood balancing. Autoimmune diseases

76 and nutrient deficiencies are the principal causes of thyroid complications [Monaco, 2003]. Thyroid

77 dysfunction is rather common in the general population, and mild or sub-clinical forms can be present in

78 more than 10% of individuals older than 80 years. Diagnosing thyroid disease in old individuals can be

79 difficult due to the non-specific clinical presentation of thyroid dysfunction and the impact of aging on the

80 test results.

81   There are different kinds of thyroid dysfunction, namely, goiter, hyperthyroidism, hypothyroidism,

82 malignant thyroid nodules, thyroiditis, etc. They are briefly summarized in the following [Monaco, 2003]:

83   • Goiter is a non-cancerous enlargement of the thyroid gland. The most common cause of goiter

84   worldwide is iodine deficiency in the diet;

85   • Hyperthyroidism occurs when the thyroid gland is overactive. It produces too much of its hormone;

86   • Hypothyroidism is the opposite of hyperthyroidism. The thyroid gland is underactive, and it cannot

87   produce enough of its hormones;

88   • Thyroid nodules are growths that form on or in the thyroid gland. The nodules can be solid or

89   fluid-filled, most are benign, but they can also be cancerous in a small percentage of cases;

90   • Thyroiditis can be considered a swelling of the thyroid.

91 Another pathology is Euthyroid, a normal thyroid hormonal functional state, but involved in initial

92 structural changes such as goiter, cold nodule, multiple nodule goiter (MNG) and cancer (Grave's Disease

93 and the like).

94   In the rest of the paper, we will specifically focus on the following general and more frequent

95 thyroid-related diseases: hypo- and hyperthyroidism, thyroid cancer, and euthyroid sick state. Other

96 thyroid-related diseases will be considered, but in a general thyroid disease category.

3 RELATED WORK

98 In this section, we summarize the existing reviews and surveys on the studied topic and we highlight the

99 differences, compared with our proposed literature review. First of all, it is important to specify that in the

literature there are currently very few systematic reviews or surveys about the use of artificial intelligence detection techniques for thyroid dysfunction.

Some studies discuss a more general context regarding both clinical and computerized thyroid dysfunction detection techniques. Parmar and Mehta [2020], for example, discuss different kinds of dysfunctions which affect the thyroid and point out the main methods and processes used currently for detecting these dysfunctions from a clinical perspective. Moreover, they illustrate the computer-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 the feature extraction method and classification approach, is used to compare the discussed techniques. A general survey is also proposed in [Razia et al., 2020] reporting an overview of various machine learning techniques in medicine.

Alternative contributions are generally focused on one particular diagnostic method [Chen et al., 2020], thyroid disease [Abdolali et al., 2020; Chen et al., 2020; Garg and Mago, 2021] or one specific task (i.e., segmentation, classification, diagnosis) [Ludwig et al., 2023]. For example, Chen et al. [2020] propose a review and categorization of thyroid gland and thyroid nodule segmentation methods according to their own theoretical bases. In particular, the review compared 28 representative papers selected in the literature and found out that the most common methods for thyroid disease detection are based on machine and deep learning techniques. Moreover, the study found that the usage of big data for training provide better segmentation performance and robustness. However, deep learning models usually require large training datasets and imply a long training time. For thyroid nodule segmentation, the most commonly adopted methods are contour and shape-based methods, which lead to satisfactory performance results. Nevertheless, they are often tested on small datasets.

Similarly, in [Abdolali et al., 2020] and in [Ludwig et al., 2023], the authors provide a systematic review of artificial intelligence applications focused on thyroid cancer diagnosis. Abdolali et al. [2020] considered and classified more than 50 papers discussing approaches for thyroid cancer detection exploiting AI algorithms. The paper also proposed future trends and challenges in the field and perspectives of computer-aided analysis to improve the efficiency of future methods for thyroid cancer diagnosis. In [Ludwig et al., 2023] and Casella et al. [2022], the particular focus is on the usefulness of AI in ultrasonography for the diagnosis and characterization of thyroid cancer. Sorrenti et al. [2022] also report an overview of the state of the art regarding AI implementation for thyroid nodule ultrasound characterization and cancer. These recent studies show an increasing interest in the last year about the adoption of AI in the thyroid nodule ultrasound context. Finally, Mendoza and Hernandez [2021] provide a short review assessing and analyzing existing data mining methods for diagnosing thyroid diseases. This study, in the limited room of a conference, does not provide any keywords or research questions, thus does not adhere to Barbara Kitchenham's guidelines [Kitchenham, 2004] for a systematic review either.

There are also recent reviews that have a specific focus on a certain type of data. Considering AI in the context of improving thyroid health, some works focus only on images, such as [Bini et al., 2021] and

[137] [Sharifi et al., 2021], and do not consider clinical data at all.

[138] Differently from all the aforementioned papers, the proposed systematic literature review aims to [139] perform a comprehensive investigation on the use of artificial intelligence approaches for the diagnosis [140] and classification of the main thyroid dysfunctions, providing the mapping of the considered approaches [141] with the used data types and the available datasets. We have considered different data types, like images [142] and clinical data, and we have included all main thyroid dysfunctions. We have filtered out works tackling [143] thyroid diseases only as an example of the application of machine learning techniques; moreover, we [144] have carefully adhered to Barbara Kitchenham's guidelines and the PRISMA selection procedure Page [145] et al. [2021] for a systematic review, clearly pointing out the investigated research questions, the formal [146] analysis and filtering process we have followed, as well as the final results and discussion.

4 RESEARCH METHOD

[148] In order to properly conduct the literature review, we mainly adopted the guidelines proposed by Kitchen-[149] ham [2004] and the PRISMA selection procedure Page et al. [2021] for a systematic review. The phases [150] performed are described in the following subsections.

4.1 Research Questions and relevant keywords

[152] The research questions we used to investigate the application of AI techniques to the detection of thyroid [153] diseases are the following:

[154] • RQ1: What are the main AI techniques used for the detection and identification of the most relevant [155] thyroid diseases?

[156] • RQ2: What datasets about thyroid diseases are used in the considered AI solutions?

[157] • RQ3: What data types are used to detect and classify thyroid diseases using the considered AI [158] techniques?

[159] The first research question aims to discriminate the various thyroid diseases tackled by different [160] AI-based detection approaches as well as their performance metrics.

[161] AI field is rapidly advancing comprising a wide range of approaches and techniques. To understand [162] the different ways in which AI is being used and developed for thyroid disease treatment, we have [163] adopted a taxonomy of AI approaches, shown in Figure 1, we used to classify and organize the various [164] methods and applications of AI we found. This provides a framework for understanding the field and [165] its potentials in the context of our systematic review. The selected categories are derived from [Witten [166] et al., 2011] and are roughly based on the main categories of AI techniques (expert systems based on rule [167] mining, decision tree-based or probabilistic-based machine learning models, artificial neural networks [168] and hybrid/composite approaches):

[169] 1. **Probabilistic Approaches**, which classify or group samples according to a certain probability [170] distribution function;
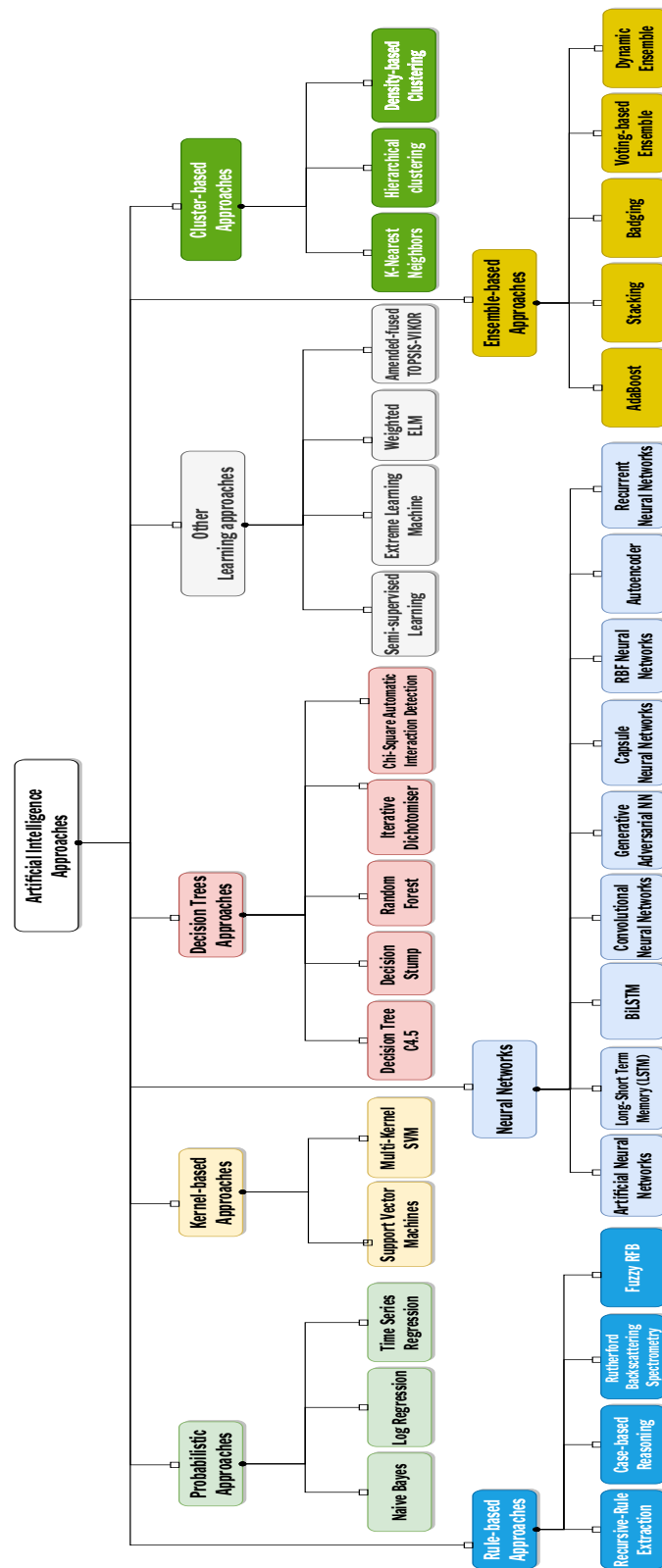
Figure 1. The considered taxonomy of Artificial Intelligence approaches for the classification of the surveyed papers.

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;

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

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 or **hybrid methodologies**.

Conversely, for what concerns 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 the considered features. In this study, we have considered as clinical data all the structured data describing the state of health of the patients as well as their sociodemographic characteristics. These data are usually managed and analyzed in a different manner with respect to the medical images and other types of data (i.e., natural language text).

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:

Table 1. Criteria for including or excluding papers in the performed search process.

Acronym	Description of the criterium
Inclusion Criteria	
IC_1	Studies published in the year 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 thyroid

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";

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.

Moreover, we have also searched the PeerJ CS database[5], to verify whether the considered topic was already regarded in the intended publication venue.

4.3 Search process and filtering criteria

We have performed the search and filtering process represented in the PRISMA diagram in Figure 2, wherein the inclusion (IC) and exclusion criteria (EC) presented in Table 1 were applied.

The selected papers fall in the January 2016 - July 2022 range (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

---

[1] https://ieeexplore.ieee.org/Xplore/home.jsp

[2] https://www.sciencedirect.com/

[3] https://link.springer.com/

[4] https://dl.acm.org/

[5] https://peerj.com/computer-science/
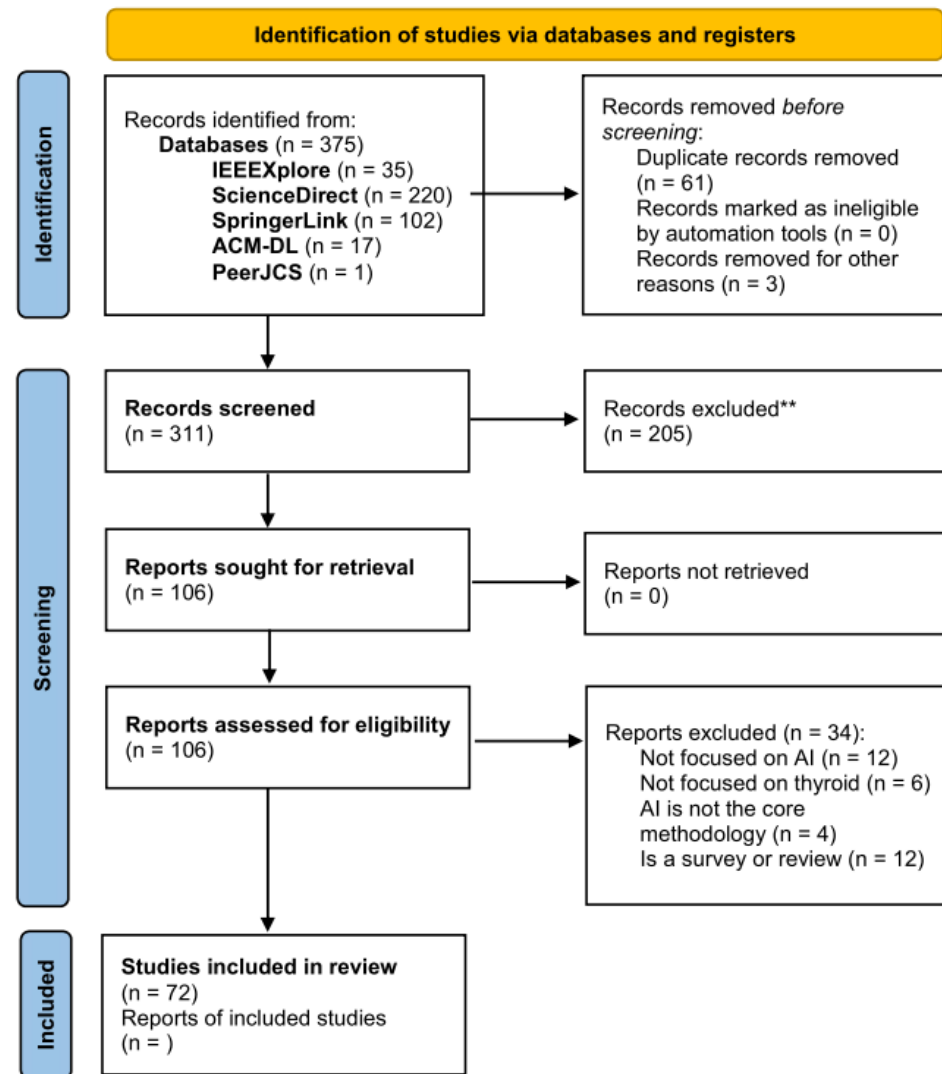
Figure 2. PRISMA diagram of the selection procedure for the analyzed papers.

queries to the aforementioned databases we got a total of 375 papers, 35 from IEEEXplore, 220 from ScienceDirect, 102 from SpringerLink, and 17 from the ACM Digital Library. We have also found one paper in the PeerJ CS database.

Then, we skimmed titles and abstracts in order to remove possible duplicated as well as 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 106 papers remained. Finally, these remaining articles were retrieved and fully read 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);

235   • specifically focused on thyroid (EC_3);

236   • an AI-based technique must be the core method of the paper to identify or classify thyroid diseases

237   (IC_3);

238   • not a survey nor a review (EC_1).

239 This final filtering step resulted in a total of 72 papers to analyze, after removing also 12 recent surveys,

240 some thereof already discussed in Section 3.

241   Figure 3 presents the distribution over the years of the 72 papers selected for the systematic review. As

242 one can see, there is a constant growth of the research on the topic over the recent years with the obvious

243 exception of 2022, given that it can only consider half of the year. The trend of the graph demonstrates an

244 increasing interest in the usage of AI-based methods to detect and identify thyroid-related dysfunctions,

245 thus corroborating our initial idea to research on this hot topic.
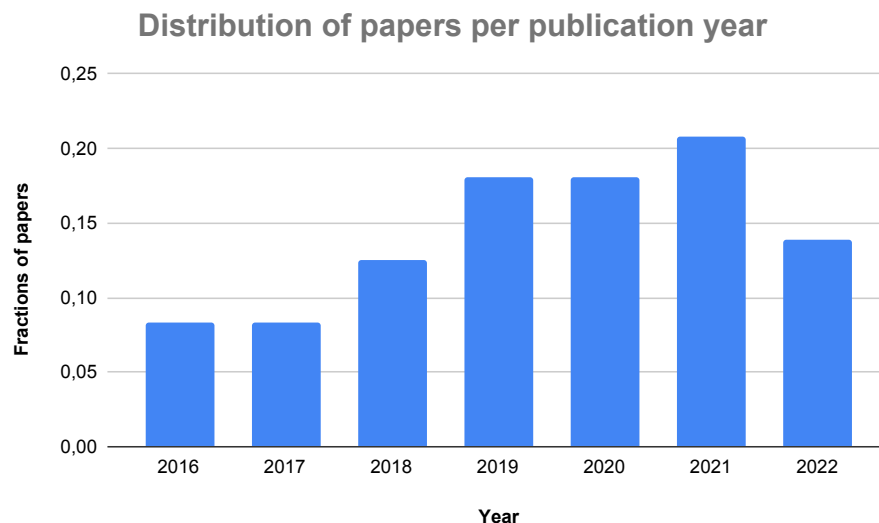
Figure 3. Distribution of the considered papers over the last years.

246   5   RESULTS

247 In this section, we discuss the main results and outcomes of the systematic literature review, by following

248 the research questions described in Section 4.

249   5.1   RQ1: What are the main AI techniques used for the classification and identification of the most

250   relevant thyroid diseases?

251 This research question regarded the identification of the various AI techniques, as well as the main tackled

252 thyroid diseases, used in the 72 papers identified after the filtering process. Figure 4 summarizes the

253 percentage of the identified papers as regards a raw classification of the AI techniques. This classification
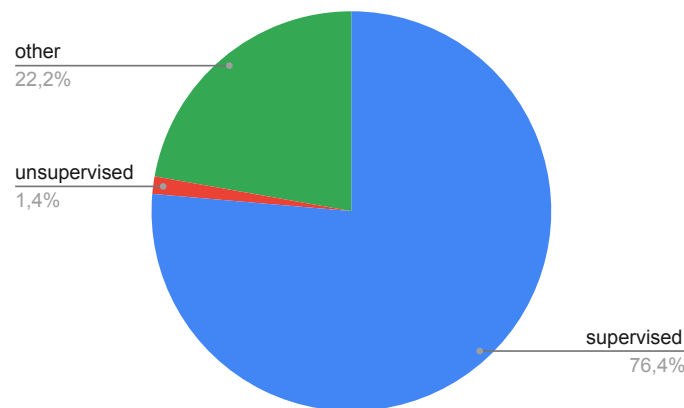
Figure 4. Pie chart of the survived papers according to a coarse classification of the AI techniques.

²⁵⁴ entails three main classes, namely, supervised, unsupervised, and other methods. The first ones regard

²⁵⁵ techniques wherein the classification is performed by exploiting labels that are already known, the second

²⁵⁶ ones try to find out a classification with no prior knowledge of any label, the third ones take advantage

²⁵⁷ of either a mix of both of the aforementioned or approaches which cannot be framed in the previous

²⁵⁸ categories.

²⁵⁹ From the figure, it can be noted that the most of the papers (76.39%) entails supervised techniques,

²⁶⁰ while unsupervised approaches are a very small fraction (1.41%). Conversely, other or hybrid methodolo-

²⁶¹ gies are quite relevant, reaching a percentage equal to 22.22%.

²⁶² In Figure 5, there is a classification of the considered AI techniques on the basis of the four main

²⁶³ thyroid disease categories we have considered. In this case, a study may face more than one disease, thus

²⁶⁴ the number of occurrences is not computed over 72 papers, but over 119 tackled diseases (36 papers face

²⁶⁵ cases of hypothyroidism, 33 papers face cases of hyperthyroidism, 11 papers regard euthyroid disease, 22

²⁶⁶ papers face cases of thyroid cancer, and 17 papers discuss generic thyroid diseases or other thyroid-related

²⁶⁷ illnesses).

²⁶⁸ In Table 2, we distribute the surveyed papers according to both the AI techniques and the just

²⁶⁹ mentioned tackled thyroid diseases. The table also groups the various AI techniques according to the raw

²⁷⁰ taxonomy depicted in Figure 1.

²⁷¹ In Figure 6, we show a classification of the considered AI taxonomy. Also in this case, a study may

²⁷² exploit more than one technique, thus the relative fraction is not computed over 72 papers, but over 140

²⁷³ used techniques, considering repetitions in each AI group (i.e., CNNs are used in 17 papers, SVM is used

²⁷⁴ in 17 papers, KNN in 10 papers). As one can see, the most used techniques to face thyroid diseases are

²⁷⁵ neural networks, building up more than one third of the overall AI-based methods (35%), followed by

²⁷⁶ Decision Trees (16.43%), kernel-based methods (12.86%), probabilistic techniques (11.43%), ensemble

²⁷⁷ techniques (8.57%), and cluster-based methods (7.14%). The other methods represent a fraction smaller

²⁷⁸ than 7%.
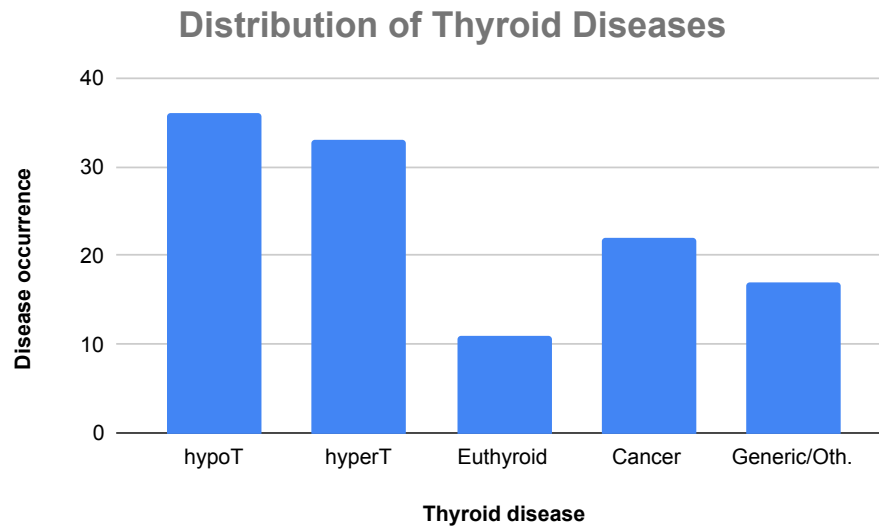
Figure 5. Absolute occurrences of the tackled thyroid diseases in the considered papers.
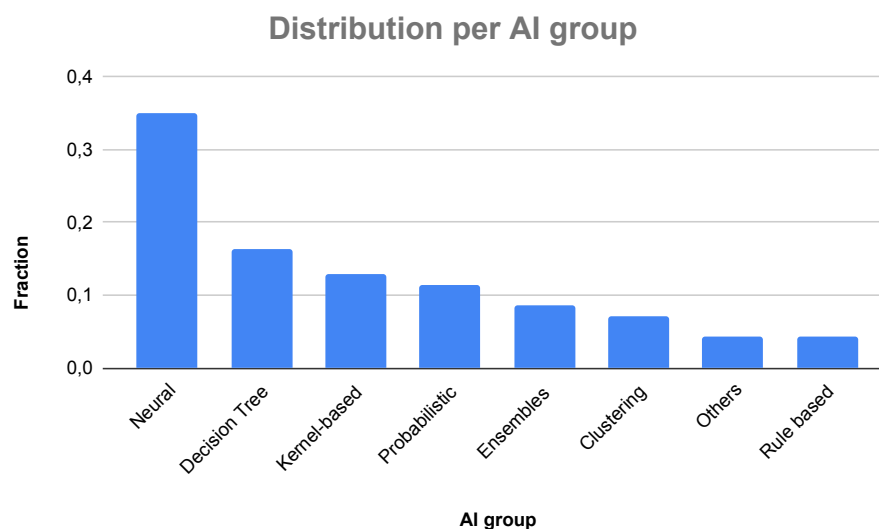
Figure 6. Fractions of the various AI groups.

Finally, Table 3 presents the average, standard deviation, 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. As one can see, average accuracy is rather high (96.84%) with a small standard deviation, whereas F1-score is high (90.39%), but with a great standard deviation (about 9%).

Table 2. Cross classification of the considered papers: according to the tackled thyroid disease and according to the various AI techniques, grouped also per macro-category.

AI group	AI technique	Hypothyroidism	Hyperthyroidism	Euthyroid disease	Thyroid Cancer	Generic/others
Probabilistic approaches	Naïve Bayes[6]	Rao and Remaka [2020]|Pasha and Mohamed [2020]|Houari et al. [2016] Chandel et al. [2016]|Duggal and Shukla [2020]|Peya et al. [2021] Riajulislam et al. [2021]|Juneja [2022]	Rao and Remaka [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]|Islam et al. [2022]	Qin et al. [2021]	Juneja [2022] Kishor and Chakraborty [2021]
	Log regression[7]	Pasha and Mohamed [2020]|Riajulislam et al. [2021]	Pasha and Mohamed [2020]		Qin et al. [2021]	Raisinghani et al. [2019]|Chanbey et al. [2021]
	Time Series Regression[8]	Chandioi et al. [2016]		Chandio et al. [2016]		
Kernel-based approaches	Support Vector Machine[9]	Duggal and Shukla [2020]|Shahid et al. [2019]|Ahmed and Soomrani [2016] Tyagi et al. [2018]|Shen et al. [2016]|Li et al. [2019] Chandel et al. [2016]|Zarin Mousavi et al. [2020]|Peya and Srinivasan [2017] Riajulislam 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. [2019] Chandel et al. [2016]|Peya and Srinivasan [2017]	Ahmed and Soomrani [2016]|Shen et al. [2016]|Li et al. [2019] Duggal and Shukla [2020]|Islam et al. [2022]	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] Kishor and Chakraborty [2021]
	multi-kernel SVM	Kumar [2020]		Kumar [2020]		
Decision Tree approaches	Decision Tree[10]	Rao and Remaka [2020]|Sidiq and Mutahar Aaqib [2019]|Tyagi et al. [2018] Hayashi [2017]|Peya et al. [2021]|Riajulislam et al. [2021] Juneja [2022]	Rao and Remaka [2020]|Sidiq and Mutahar Aaqib [2019]|Tyagi et al. [2018] Hayashi [2017]|Peya et al. [2021]|Juneja [2022]	Islam et al. [2022]	Hao et al. [2018]	Tyagi et al. [2018]|Raisinghani et al. [2019]|Kaur et al. [2019] Chanbey et al. [2021]|Juneja [2022]|Kishor and Chakraborty [2021]
	Decision stump[11]		Jha et al. [2018]			
	Random Forest[12]	Sidiq and Mutahar Aaqib [2019]|Shahid et al. [2019]|Duggal and Shukla [2020] Riajulislam et al. [2021]|Juneja [2022]	Sidiq and Mutahar Aaqib [2019]|Shahid et al. [2019]|Inthos et al. [2017] Duggal and Shukla [2020]|Juneja [2022]	Duggal and Shukla [2020]|Islam et al. [2022]	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]
	Iterative Dichotomizer 3[13]	Zarin Mousavi et al. [2020]	Zarin Mousavi et al. [2020]			
	Chi-squared Automatic Interaction Detection[14]	Zarin Mousavi et al. [2020]	Zarin Mousavi et al. [2020]			
	Recursive-Rule extraction[15]	Hayashi [2017]	Hayashi [2017]			
Rule-based approaches	Case-based reasoning	Bentaiba-Lagrid et al. [2020]	Bentaiba-Lagrid et al. [2020]			
	Rutherford Backscattering Spectrometry[16]		Imbus et al. [2017]			
	Fuzzy RBS	Assad Sajadi et al. [2019]|Kumari and Sharma [2019]	Kumari and Sharma [2019]			
Neural Network	Artificial neural network[17]	Sidiq and Mutahar Aaqib [2019]|Mahatkar 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]|Mahatkar and Gaikwad [2017] Tyagi et al. [2018]|Houari et al. [2016]|Vivar et al. [2020]	Islam et al. [2022]	Ahmed et al. [2022]|Cordes et al. [2021]|Jin et al. [2021]	Tyagi et al. [2018]|Raisinghani et al. [2019] Kaur et al. [2019]|Santos et al. [2019]|Kishor and Chakraborty [2021]
	Bidirectional Long Short-Term Memory - Long/Short Term Memory (LSTM)[18]	Yue et al. [2020]	Yue et al. [2020]|Li et al. [2020]			Chai [2020]
	Convolutional Neural Network[19]	Yue et al. [2020]|Amarthi et al. [2022]|Khan [2021]	Yue et al. [2020]|Amarthi et al. [2022]		Yin et al. [2019]|Yi et al. [2017]|Lyu and Haque [2018] Li et al. [2019a]|Moran et al. [2018]|Avanthi et al. [2022] Chu et al. [2021]|Liu et al. [2021]|Santillan et al. [2021] Song et al. [2022]	Pouded et al. [2019]|Guo and Du [2019]|Avanthi et al. [2022] Pi et al. [2022]|Yang et al. [2021]
	Generative Adversarial Network[20]	Zhang et al. [2020]	Zhang et al. [2020]		Shi et al. [2020]|Zhao et al. [2022]	
	Capsule Network[21]				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]	
	Recurrent Neural Network[22]				Santillan et al. [2021]	
	MultiLayer Perceptron[23] BackPropagation	Yue et al. [2020]|Zarin Mousavi et al. [2020]|Peya and Srinivasan [2017] Hosseinzadeh et al. [2021]|Juneja [2022]	Yue et al. [2020]|Peya and Srinivasan [2017]|Hosseinzadeh et al. [2021] Juneja [2022]		Qin et al. [2021]	Juneja [2022]
Cluster-based approaches	K-Nearest Neighbors Algorithm[24]	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]	Islam et al. [2022]	Qin et al. [2021]	Tyagi et al. [2018]|Kaur et al. [2019]|Chanbey 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]	Islam et al. [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]|Yadav and Pal [2018]	Islam et al. [2022]	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					
	Weighted Extreme Learning Machine[25]					Priya and Manavalan [2018]
	Extreme Learning Machine[26]	Peya and Srinivasan [2017]|Ma et al. [2018]|Juneja [2022]	Peya and Srinivasan [2017]|Ma et al. [2018]|Juneja [2022]			Juneja [2022]
	Amended fused TOPSIS-VIKOR for classification[27]	Baccour [2018]	Baccour [2018]			

Table 3. Statistics of the metrics of the surveyed AI techniques for thyroid disease classification.

	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Average	96.84 ± 1.91	94.96 ± 6.32	90.21 ± 11.90	90.39 ± 9.02
Maximum	100.00	100.00	100.00	99.00
Minimum	93.00	85.00	69.00	74.00

In the following subsections, we briefly summarize each surveyed paper according to the AI group presented in Table 2. In case more than one AI group is used in a paper, this is described in the relative first section of appearance.

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 detection by exploiting both a Random Forest-based method and a Gain Ratio technique. In the end, the prediction is performed by comparing also different machine learning techniques, namely K-Nearest-Neighbor, Logistic Regression, and Naive Bayes. Similarly, Houari et al. [2016] try to reduce redundant dimensions by exploiting Copulas and LU-decomposition techniques. They 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. 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 classifiers, whose performance is compared with Naïve 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, Naïve Bayes, adaptive boosting, Random Forest, artificial neural networks, and K-nearest neighbor. They find out that Random Forest is the best performing algorithm to predict fatal diseases about thyroid. Authors in [Qin et al., 2021] try to study papillary thyroid cancer through the analysis of magnetic resonance imaging radiomics by exploiting eight classifiers (logistic regression, bagging, Random Forests, extremely randomized trees, support vector machines, Naïve 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 disease prediction model is proposed exploiting 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 decision tree resulted the

313 most accurate algorithm, with a 99.7% of accuracy.

314 Riajuliislam et al. [2021] try to predict early stage hypothyroidism. They apply three different

315 feature selection procedures: Principal Component Analysis (PCA), Recursive Feature Elimination (RFE),

316 univariate features selection (UFS). Moreover, they use different classification algorithms (i.e., support

317 vector machine, decision tree, random forest, logistic regression, and Naive Bayes). RFE applied to

318 the UCI thyroid dataset allowed the authors to achieve a constant 99% accuracy value over the various

319 considered classification algorithms.

320 In [Raisinghani et al., 2019], the authors compare different machine learning approaches, (Logistic

321 regression, Decision trees, Random forest, Support Vector Machine) to develop predictive models to

322 detect a generic thyroid disease.

323 The contribution in [Chandio et al., 2016] exploits Time Series Regression to create an intelligent

324 system for thyroid disease visualization. This allows a careful surveillance of the thyroid disease, with a

325 particular focus on hypothyroidism, hyperthyroidism, and euthyroid disease.

326 The study in [Chaubey et al., 2021] compares logistic regression, decision tree, and kNN on the UCI

327 knowledge discovery database for thyroid diseases detection. The best result of accuracy is obtained by

328 applying the kNN classifier (96.87%).

329 Finally, Islam et al. [2022] try to detect euthyroid disease by leveraging eleven machine learning

330 classifiers, namely, Gaussian Naive Bayes, ANN, CatBoost, XGBoost, Random Forest, LightGBM,

331 Decision Tree, Extra-trees, SVC, and KNN, and applying also feature selection to the sick-euthyroid

332 dataset. The best performing classifier resulted to be ANN with a maximum accuracy equal to 95.87%.

333 5.1.2 Kernel-based approaches

334 Duggal and Shukla [2020] perform feature selection and extraction before applying Naive Bayes, Support

335 Vector Machine, and Random Forest to identify hypothyroidism, hyperthyroidism, and euthyroid disease.

336 The authors of [Shahid et al., 2019] compare Random Forest, Support Vector Machine, and K-Nearest

337 Neighbours, on the UCI thyroid dataset, to discover the best performing algorithm, resulted to be Random

338 Forest, in detecting hypo- and hyperthyroidism.

339 The contribution in [Ahmed and Soomrani, 2016] provides a framework, named Thyroid Disease

340 Types Diagnostics (TDTD), aiming at making a diagnosis of various thyroid diseases in a very structured

341 and transparent manner and exploiting binary and multi-SVM algorithms, as well as Bayesian isotonic

342 regression for missing values. In [Tyagi et al., 2018], the authors present again the results in detecting

343 hypo- and hyperthyroidism (by using the UCI dataset) through different machine learning techniques,

344 such as decision trees, artificial neural networks, support vector machines, and k-nearest neighbors.

345 Kumar [2020] introduces a novel mutliclass SVM approach to detect four types of subjects, i.e.,

346 people affected by hypothyroidism, hyperthyroidism, euthyroidism disease and healthy euthyroid, on the

347 already mentioned UCI dataset. Shen et al. [2016] propose a novel scheme to optimize the parameters of

348 SVM by means of the fly optimization algorithm. The novel scheme is compared with other optimization

349 algorithms and its efficiency is tested on four datasets, including the UCI one about thyroid.

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, with the aim of detecting malignant nodules by means of Random Forest and SVM.

Similarly, the authors of the study in [Raghavendra et al., 2017] try to propose 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.

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

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

5.1.3 Decision Tree approaches

Hayashi [2017] 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 J48 graft in terms of rule definition in the IF-THEN form 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 sub-intervals. The used dataset is made of clinical-pathological data, obtained from the First Hospital of Jilin University.

Jha et al. [2018] present a hybrid algorithm for healthcare data mining by using Decision Stump (DS), StackingC (SC), and voting methods to tackle the hyperthyroidism issue. 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, Zarin Mousavi et al. [2020] apply computational methods based on decision trees, like ID3 and CHAID, SVM and multi-layer perceptrons, enriched with bagging and boosting techniques, for 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 systems 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 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, by 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 an 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] propose 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.

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] exploit again 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. In [Shi et al., 2020], the authors integrate domain knowledge, extracted from standardized terminology, and deep learning (Auxiliary Classifier Generative Adversarial Network) into a synthetic medical image augmentation technique 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%; other results and the optimization process are not described in detail.

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. The latter only focus on hypothyroidism and achieve 98% of accuracy. Liu et al. [2021] 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%. The latter employ a feature-enhanced dual branch convolutional neural network on 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, by 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%.

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 euthyroid disease.

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 [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 hyperthyroidism). Finally, in [Baccour,

496 2018], a new classification system for both hypo- and hyperthyroidism and based on fused VIKOR and

497 TOPSIS is proposed.

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

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

500 In Figure 7, we show the availability of the datasets used in the papers we considered in this systematic

501 review. As one can see, the number of private or non-available datasets is higher (55.42%) compared with

502 the number of publicly available ones.
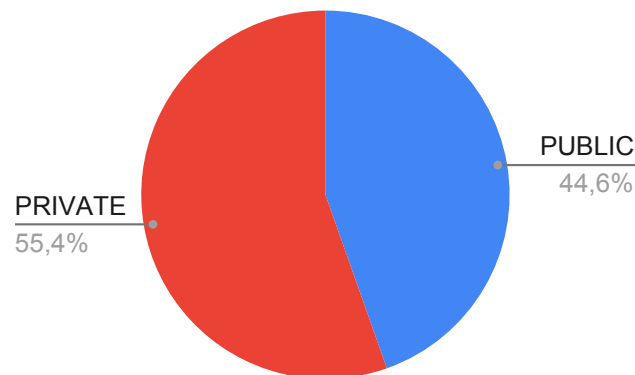
Figure 7. Availability of the datasets in the surveyed papers.

503 Looking the public datasets, we observed that the most used dataset is the UCI one[28], exploited 27

504 times (Duggal and Shukla [2020], Shahid et al. [2019], Pan et al. [2016], Pavya and Srinivasan [2017],

505 Mahurkar and Gaikwad [2017], Ahmed and Soomrani [2016], Tyagi et al. [2018], Kumar [2020], Pasha

506 and Mohamed [2020], Shen et al. [2016], Bentaiba-Lagrid et al. [2020], Raisinghani et al. [2019], Vivar

507 et al. [2020], Li et al. [2019b], Ma et al. [2018], Kour et al. [2020],Khan [2021], Priyadharsini and Sasikala

508 [2022], Peya et al. [2021], Chaubey et al. [2021],Hosseinzadeh et al. [2021], Juneja [2022], Kishor and

509 Chakraborty [2021], Islam et al. [2022], Saktheeswari and Balasubramanian [2021], Chandel et al. [2016],

510 Priya and Manavalan [2018]). The UCI dataset is characterized by 7,200 instances and 21 categorical and

511 real attributes. It is quite old since it was built in 1987.

512 The study proposed in Rao and Renuka [2020] use a dataset available on Kaggle Machine Learning

513 Website[29]. This dataset consists of the 3,162 instances, including basic patient information details and

514 clinical history. The considered 27 attribute values are boolean or continuous.

515 A public dataset[30] is also used in Moran et al. [2018]. This dataset contains the infrared image,

516 clinical, and histopathological data collected from patients with thyroid nodules at the Outpatient Clinic

517 of Endocrinology and Surgery of the Hospital Universitario Antonio Pedro - HUAP of the Fluminense

518 Federal University (UFF). The number of instances considered in the study is 92 with 20x20 featuring

---

[28]https://archive.ics.uci.edu/ml/datasets/Thyroid+Disease

[29]https://www.kaggle.com/datasets/kumar012/hypothyroid

[30]http://visual.ic.uff.br/en/thyroid/

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

	Instances	**Features**
Average	5641	25674
Maximum	92062	711680
Minimum	92	4

519 images.

520   Finally, in Table 4, we present some statistics about the main characteristics of the datasets used in the

521 surveyed papers. They show that the average number of instances per dataset is about 5,641 samples,

522 while the average number of features is almost 26,000, but in this case we have to consider that for image

523 datasets the number of features was considered equal to the number of pixels, thus making the statistics

524 increase very much.

525 **5.3 RQ3: What data types are used to detect and classify thyroid diseases using the considered AI**

526   **techniques?**

527 In this section, we analyze the types of features considered in the surveyed papers.

528   In Figure 8, we show a pie chart to highlight the different percentages of data used in the surveyed

529 papers. As it can be easily inferred, the majority of the considered features (51.85%) is made of clinical

530 data, i.e., data mainly coming from the blood analysis, while 27.16% of the surveyed papers use images

531 of the thyroid, and only a small percentage (12.35%) exploit both types of data.
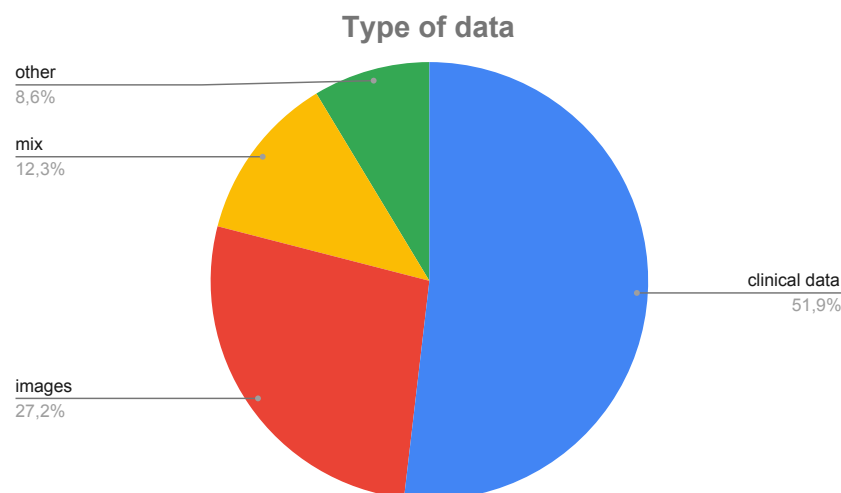
Figure 8. Type of data used in the considered datasets.

532   Table 5 shows the occurrences of the ten most used features in the surveyed articles. As it is evident,

533 the most used feature is the thyroid-stimulating hormone (TSH) with 39 occurrences, followed by

534 triiodothyronine (T3) hormone, the age, and the gender of the subject. Other very used features regard the

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

Feature	Occurrences
TSH[31]	39
T3[32]	28
Age	28
Gender	27
Pixel value	19
TT4[33]	19
FTI [34]	18
T4U[35]	18
Pregnancy	14
T4[36]	14

pixel value (20), the only feature used whenever image datasets are employed, as well as the thyroxine, also called T4, FTI (Free T4 Index) and T4U (T4 Uptake). Pregnancy status is the least important feature in the top ten context, and can be obviously used only in female patients.

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 papers focuses 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 detection of thyroid dysfunctions and diseases. Moreover, they ensure a possible solution when there is a limited labeling budget [Yoon et al., 2022]. In some cases, effectively, labels are particularly tedious to obtain, and labeled samples are not available during the training step. Finally, even when labeled data are available, there could be biases in the way samples are labeled, causing distribution differences.

We also observed that the greatest number (54 papers) of the surveyed studies is focused on offering a solution to the issue of disease detection, while a very poor number of papers faces the topic of disease prediction and monitoring. The main reason, according to our opinion, is the absence of useful datasets for this scope.

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 the use of AI techniques in the detection of thyroid dysfunction.

The average performance of the surveyed AI techniques in detecting thyroid-related diseases can be further improved (Table 2 shows an average F1-score of 90.39%). Given the 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 topic of future research papers.

Moreover, there are few datasets publicly available, and among these, the most used one is the UCI dataset, which is quite old (1987). The need for novel, public, and updated datasets is, thus, compelling, together with the application of proper feature selection and feature reduction techniques, which are very poorly investigated in the surveyed papers.

However, based on the declarations of the authors of the analyzed studies, we observed that their main limitation is in the quality and dimensions of the adopted datasets. Some authors highlight the necessity to increase the adopted dataset Poudel et al. [2019]; Yin et al. [2019]; Moran et al. [2018]; Kaur et al. [2019]; Ma et al. [2018]; Qin et al. [2021]; Shen et al. [2021]. In other studies, authors identify as a limit the reduced heterogeneity and balancing of the adopted dataset Pan et al. [2016]; Yi et al. [2017]; Pi et al. [2022]; Vivar et al. [2020]. Moreover, in the surveyed datasets, the usage of clinical data is prevalent; thus, also in view of the greater usage of various deep learning techniques, greater development of public image datasets to be used with proper deep convolutional neural networks should be encouraged.

Finally, even if in all considered studies there is the adoption of real data (the data are collected by medical institutions from patients), the proposed approaches are almost never applied in a real context. This highlights that there is still a lot of resistance from medical institutions to use AI solutions to support their medical activities.

7 CONCLUSIONS

In this paper, we have performed a detailed systematic review about the application of artificial intelligence techniques to the study, analysis, and detection 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 72 analyzed papers.

In the end, we have presented and discussed in depth the results of the systematic review: the main AI techniques used for the classification and identification of the most relevant thyroid diseases (RQ1), the datasets, and data types used in the considered AI solutions (RQ2, RQ3).

The obtained results show that supervised methods are very widespread (they are used in 76,39% of the analyzed papers) while unsupervised approaches are a very small fraction (1.39%).

This, in our opinion, is due to the greater interest of the medical staff in classification tasks and the presence of large labeled datasets. However, the interest about unsupervised and hybrid approaches should be also encouraged because of their capability of learning the data and classifying them without any labels.

Moreover, the most used AI technique to face thyroid diseases is neural networks, building up

more than one-third of the overall AI-based methods (35%).The great interest about neural networks is motivated by the advantages of their application in medical domain. However, their ability to learn and model complex relationships is really important in the considered domain where many of the relationships between inputs and outputs are non-linear and complex.

Finally, a large number of papers study hypothyroidism and hyperthyroidism (30,25% and 27,73%). The reason for this is that hypothyroidism and hyperthyroidism are the two most common thyroid pathologies according to the most recent studies Muñoz-Ortiz et al. [2020].

Looking at the adopted datasets, we observed that there is a high number of not available datasets (55,42%) and this limits the possibility to extend the existing research. However, the most used dataset (27 times in the considered papers) is the UCI one, which is quite old (it was published in 1987). Finally, we observed that the majority of the considered features (51.85%) are made of clinical data, while 27.16% of the surveyed papers use images of the thyroid, and only a small percentage (12.35%) exploit both types of data. This, according to our opinion is due to the reduced availability of datasets containing both images and clinical data.

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

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., Iammarino, M., Macchia, P. E., Nettore, I. C., and Verdone, C. (2021a). Thyroid disease treatment prediction with machine learning approaches. *Procedia Computer Science*, 192:1031–1040. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 25th International Conference KES2021.

Aversano, L., Bernardi, M. L., Cimitile, M., and Pecori, R. (2021b). 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).

Casella, C., Xue, Y., Zhou, Y., Wang, T., Chen, H., Wu, L., Ling, H., Wang, H., Qiu, L., Ye, D., and Wang, B. (2022). Accuracy of ultrasound diagnosis of thyroid nodules based on artificial intelligence-assisted diagnostic technology: A systematic review and meta-analysis. *International Journal of Endocrinology*, 2022:9492056.

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*,

664   185:105329.

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

667   Cordes, M., Götz, T. I., Lang, E. W., Coerper, S., Kuwert, T., and Schmidkonz, C. (2021). Advanced
668   thyroid carcinomas: neural network analysis of ultrasonographic characteristics. *Thyroid Research*,
669   14(1):16.

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

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

675   Guo, M. and Du, Y. (2019). Classification of thyroid ultrasound standard plane images using resnet-
676   18 networks. In *2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and*
677   *Identification (ASID)*, pages 324–328.

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

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

683   Hosseinzadeh, M., Ahmed, O. H., Ghafour, M. Y., Safara, F., hama, H. k., Ali, S., Vo, B., and Chiang, H.-S.
684   (2021). A multiple multilayer perceptron neural network with an adaptive learning algorithm for thyroid
685   disease diagnosis in the internet of medical things. *The Journal of Supercomputing*, 77(4):3616–3637.

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

688   Imbus, J. R., Randle, R. W., Pitt, S. C., Sippel, R. S., and Schneider, D. F. (2017). Machine learning to
689   identify multigland disease in primary hyperparathyroidism. *Journal of Surgical Research*, 219:173–
690   179.

691   Islam, S., Haque, M., Miah, M., Sarwar, T., and Nugraha, R. (2022). Application of machine learning
692   algorithms to predict the thyroid disease risk: an experimental comparative study. *PeerJ Computer*
693   *Science*.

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

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

700   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.

Keestra, S., Högqvist Tabor, V., and Alvergne, A. (2020). Reinterpreting patterns of variation in human thyroid function: An evolutionary ecology perspective. *Evolution, Medicine, and Public Health*, 9(1):93–112.

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.

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.

Kour, H., Manhas, J., and Sharma, V. (2020). Usage and implementation of neuro-fuzzy systems for classification and prediction in the diagnosis of different types of medical disorders: a decade review. *Artificial Intelligence Review*, 53(7):4651–4706.

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.

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, Z., Zhong, S., Liu, Q., Xie, C., Dai, Y., Peng, C., Chen, X., and Zou, R. (2021). 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.

Ludwig, M., Ludwig, B., Mikuła, A., Biernat, S., Rudnicki, J., and Kaliszewski, K. (2023). The use of artificial intelligence in the diagnosis and classification of thyroid nodules: An update. *Cancers*, 15(3).

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.

Muñoz-Ortiz, J., Sierra-Cote, M., Zapata-Bravo, E., Valenzuela-Vallejo, L., Marin-Noriega, M., Uribe-Reina, P., Terreros-Dorado, J., Gómez-Suarez, M., Arteaga-Rivera, K., and De-La-Torre, A. (2020). Prevalence of hyperthyroidism, hypothyroidism, and euthyroidism in thyroid eye disease: A systematic review of the literature. *Systematic Reviews*, 9(1).

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., and Moher, D. (2021). The prisma 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372.

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.

775 Parmar, B. S. and Mehta, M. A. (2020). Computer-aided diagnosis of thyroid dysfunction: A survey. In
776   Bellatreche, L., Goyal, V., Fujita, H., Mondal, A., and Reddy, P. K., editors, *Big Data Analytics*, pages
777   164–189, Cham. Springer International Publishing.

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

781 Pavya, K. and Srinivasan, B. (2017). Feature selection algorithms to improve thyroid disease diagnosis. In
782   *2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT)*,
783   pages 1–5.

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

787 Pi, Y., Yang, P., Wei, J., Zhao, Z., Cai, H., and Yi, Z. (2022). Fusing deep and handcrafted features
788   for intelligent recognition of uptake patterns on thyroid scintigraphy. *Knowledge-Based Systems*,
789   236:107531.

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

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

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

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

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

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

811 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.

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).

849   Knowledge-guided synthetic medical image adversarial augmentation for ultrasonography thyroid
850   nodule classification. *Computer Methods and Programs in Biomedicine*, 196:105611.

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

853 Song, R., Zhu, C., Zhang, L., Zhang, T., Luo, Y., Liu, J., and Yang, J. (2022). Dual-branch network
854   via pseudo-label training for thyroid nodule detection in ultrasound image. *Applied Intelligence*,
855   52(10):11738–11754.

856 Sorrenti, S., Dolcetti, V., Radzina, M., Bellini, M. I., Frezza, F., Munir, K., Grani, G., Durante, C.,
857   D'Andrea, V., David, E., Calò, P. G., Lori, E., and Cantisani, V. (2022). Artificial intelligence for
858   thyroid nodule characterization: Where are we standing? *Cancers*, 14(14).

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

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

867 Witten, I. H., Frank, E., and Hall, M. A. (2011). *Data Mining, Practical Machine Learning Tools*. Elsevier.

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

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

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

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

879 Yoon, J., Sohn, K., Li, C.-L., Arik, S. O., and Pfister, T. (2022). Spade: Semi-supervised anomaly
880   detection under distribution mismatch.

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

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

886   *Medicine Unlocked*, 18:100281.

887   Zhang, H., Huang, Z., and Lv, Z. (2020). Medical image synthetic data augmentation using gan. In

888   *Proceedings of the 4th International Conference on Computer Science and Application Engineering*,

889   CSAE 2020, New York, NY, USA. Association for Computing Machinery.

890   Zhao, J., Zhou, X., Shi, G., Xiao, N., Song, K., Zhao, J., Hao, R., and Li, K. (2022). Semantic consistency

891   generative adversarial network for cross-modality domain adaptation in ultrasound thyroid nodule

892   classification. *Applied Intelligence*, 52(9):10369–10383.