**Exploring the Challenges of Diagnosing Thyroid Disease with Imbalanced Data and Machine learning: A Systematic Literature Review**

**Dhekre Saber Saleh1, Mohd Shahizan Othman2**

1,2 Faculty of Computing, Universiti Teknologi Malaysia, Johor, Malaysia

\*Corresponding author: 1[saber.saleh@graduate.utm.my](mailto:saber.saleh@graduate.utm.my) (+601111738705), 2shahizan@utm.my

\*ORCID ID: 1https://orcid.org/0000-0002-4915-441X, 2https://orcid.org/0000-0003-4261-1873

**Abstract:**

Thyroid disease is a common disease affecting millions worldwide. Early diagnosis and treatment of thyroid disease can help prevent more serious complications and improve long-term health outcomes. However, thyroid disease diagnosis can be challenging due to its variable symptoms and limited diagnostic tests. By processing enormous amounts of data and seeing trends that may not be immediately evident to human doctors, Machine Learning (ML) algorithms may be capable of increasing the accuracy with which thyroid disease is diagnosed. A systematic literature review (SLR) strategy is used in this study to give a comprehensive overview of the existing literature on forecasting data on thyroid disease diagnosed using ML. This study includes 168 articles published between 2013 and 2022, gathered from high-quality journals and applied meta-analysis. The thyroid disease diagnoses (TDD) category, techniques, applications, and solutions were among the many elements considered and researched when reviewing the 41 articles of cited literature used in this research. According to our SLR, the current technique's actual application and efficacy are constrained by several outstanding issues associated with imbalance. In TDD, the technique of ML increases data-driven decision-making. In the Meta-analysis, 168 documents have been processed, and 41 documents on TDD are included for observation analysis. The limits of ML that are discussed in the discussion sections may guide the direction of future research. Regardless, this study predicts that ML-based thyroid disease detection with imbalanced data and other novel approaches may reveal numerous unrealized possibilities in the near future. This study seeks to discover the most recent ML-based and data-driven developments and strategies in diagnosing thyroid disease using imbalanced.

**Keywords**: Classification, Deep learning, Imbalanced data, Machine learning, Thyroid disease.

**تشخيص أمراض الغدة الدرقية على أساس التعلم الآلي: مراجعة منهجية للأدبيات**

**ذكري صابر صالح ,محمد شهيزان عثمان**

**كلية الحاسبات ، جامعة التكنولوجيا ماليزيا ، جوهور ، ماليزيا**

**الخلاصة**

مرض الغدة الدرقية مرض شائع يصيب الملايين في جميع أنحاء العالم. ومن الممكن أن يساعد التشخيص والعلاج المبكر لمرض الغدة الدرقية لمنع حدوث المضاعفات أكثر خطورة وتحسين النتائج الصحية على المدى الطويل. ومع ذلك ، يمكن أن يكون تشخيص مرض الغدة الدرقية صعبًا بسبب أعراضه المتغيرة والاختبارات التشخيصية المحدودة. ومن خلال معالجة كميات هائلة من البيانات ورؤية الاتجاهات التي قد لا تكون واضحة على الفور للأطباء البشريين ، قد تكون خوارزميات التعلم الآلي قادرة على زيادة الدقة في تشخيص مرض الغدة الدرقية. تُستخدم استراتيجية مراجعة منهجية للادب في هذه الدراسة لإعطاء نظرة عامة شاملة عن الأدبيات الموجودة حول التنبؤ بالبيانات المتعلقة بأمراض الغدة الدرقية التي تم تشخيصها باستخدام التعلم الالي .وتتضمن هذه الدراسة 168 مقالة منشورة بين عامي 2013 و 2022 ، تم جمعها من المجلات عالية الجودة والتحليل التلوي التطبيقي. كانت فئة تشخيص أمراض الغدة الدرقية ، والتقنيات ، والتطبيقات ، والحلول من بين العديد من العناصر التي تم النظر فيها والبحث عنها عند مراجعة 41 مقالة من الأدبيات المذكورة المستخدمة في هذا البحث. استنادًا إلى نتائج منهجية للأدب الخاص بنا ، فإن التطبيق الفعلي للتقنية الحالية وفعاليتها مقيدان بالعديد من المشكلات المعلقة المرتبطة بعدم التوازن. في تشخيص امراض الغدة الدرقية، تزيد تقنية التعليم الألي من اتخاذ القرار المستند إلى البيانات. في التحليل التلوي ،كما تمت معالجة 168 وثيقة ، و تضمين 41 وثيقة عن تشخيص امراض الغدة الدرقية لتحليل المراقبة. قد توجه حدود التعلم الالي التي تمت مناقشتها في أقسام المختصة اتجاه البحث المستقبلي. يتنبأ هذا التحليل بأن اكتشاف مرض الغدة الدرقية المستند إلى التعلم الآلي مع البيانات غير المتوازنة وغيرها من الأساليب الجديدة قد يكشف عن العديد من الاحتمالات غير المحققة في المستقبل. تسعى هذه الدراسة إلى اكتشاف أحدث التطورات والاستراتيجيات المستندة إلى التعلم الآلي والمستندة إلى البيانات في تشخيص أمراض الغدة الدرقية باستخدام غير متوازن.

**الكلمات المفتاحية:** أمراض الغدة الدرقية، التعلم الآلي ، التعلم العميق ، البيانات غير المتوازنة ، التصنيف.

**Introduction:**

The thyroid gland, which produces hormones that regulate metabolism, is affected by thyroid disease. There are two main categories of thyroid disease: hypothyroidism, characterized by a decrease in thyroid hormones and symptoms such as weight gain, fatigue, and constipation, and hyperthyroidism, characterized by an excess of thyroid hormones and symptoms such as irritability, weight loss, and tremors. Causes of thyroid disease include genetics, autoimmune disorders, and radiation exposure, according to the Mayo Clinic1. Additionally, infertility in women may be caused by thyroid gland diseases such as hypothyroidism, hyperthyroidism, and other thyroid gland disorders2.

Thyroid disease may also be caused by certain medications or issues with the pituitary gland, which regulates the thyroid. Approximately 400 million individuals are affected by thyroid diseases annually, making it a significant source of disability, as noted by the World Health Organization (WHO). Despite this, the WHO does not keep track of the number of deaths resulting from thyroid cancer3. Technology can enhance the delivery of healthcare services and bolster the health infrastructure4. However, early detection and prompt thyroid treatment can reduce fatalities5. Identifying thyroid diseases can be achieved through various means, including physical examination, blood tests, ultrasound, thyroid function tests, and thyroid biopsy6. Diagnosing thyroid disease can be challenging for healthcare providers due to symptoms similar to other illnesses, varying symptoms, limited access to specialized care, and limited diagnostic tests6.

With the advancement of machine learning in healthcare, many experts consider Thyroid Disease Diagnosis Based on Machine Learning (TDDBML) a viable option. Machine learning improves the accuracy and efficiency of disease diagnosis, and algorithms can scan large amounts of data and recognize patterns that doctors may overlook. ML algorithms can evaluate electronic health data and patient monitoring devices to identify early indications of disease7; algorithms can scan high amounts of data and recognize patterns that doctors may miss. Improving patient care and reducing wait times. The prevalence of thyroid disease and its significant impact on public health have led to the exploration of using ML for its diagnosis8.

Several investigations have suggested using ML algorithms, including Support Vector Classifier(SVC), Artificial Neural Networks (ANN), Naive Bayes, Random Forest, and K-Nearest Neighbors, for diagnosing thyroid diseases using various datasets. For instance, S Islam et al. found that the ANN classifier achieved a 96% accuracy rate in predicting thyroid diseases9. In addition, Pluciennik et al. combined molecular and clinical data using SVC (linear kernel) to improve accuracy, while Shyamala Devi et al. used multiple ML techniques with a 99% accuracy rate for predicting hypothyroid disease10. Guleria et al. achieved 100% accuracy in the early prediction of hypothyroidism using ANN11.

One of the possible limitations of ML and deep learning (DL)-based solutions is that they frequently involve sophisticated algorithms that require a large amount of data to train. This makes it hard for doctors to evaluate the algorithm's diagnosis and raises bias and reliability concerns12. For instance, DLs have numerous invisible layers, but it is not always easy to tell what role each plays in the model's predictions13. Another potential difficulty is that ML algorithms tend to support the majority class in their results. The term "majority class" refers to a dataset in which one category leads the others in total value14. It is essential for researchers and healthcare providers to carefully consider these issues when developing and using ML-based models to predict thyroid disease to ensure that they are unbiased.

Table 1summarizes the review studies that focus on the application of ML techniques in TDD and highlights the limited use of SLR compared to the focus on ML techniques. For instance, the study by K Lee et al. presents an SLR with a drawback: the machine learning methods applied to vary with the data used for thyroid disease diagnosis15. Nevertheless, since most datasets used for thyroid disease diagnosis are imbalanced, evaluating the performance of ML on such data is crucial. The study lacked a clear specification of the periods involved16. With the increasing popularity of ML-based diagnosis, applying SLR with meta-analysis is expected to address the gaps in existing studies.

The increasing number of studies in Thyroid Disease Diagnosis Based on Machine Learning (TDDBML) highlights the need for a systematic review of existing knowledge. An SLR was conducted using Scopus and WoS databases, resulting in the analysis of 168 papers, with further examination of 41 of them. The objective of the metadata analysis was to identify leading academic institutions, critical research areas, and high-quality sources. A comprehensive review of 41 publications was conducted to address the following inquiries: What are the existing DL and ML-based approaches for diagnosing thyroid disease? What are the current techniques for dealing with datasets with an imbalanced class ratio?

The SLR aims to supply as a resource for researchers by summarizing the latest methods and developments in the field and identifying gaps in knowledge that may be addressed by creating a more advanced TDDBML model. The structure of the remaining article is a methodology of the systematic literature review, briefly described in Section 2. The results and analysis are presented in Section 3, the findings are summarized in Section 4, and the conclusion is provided in Section 5.

Table Related research for TDDBML

| Paper title | Date Range | Study focus | Algorithm | Imbalance challenge | Meta-analysis | Content analysis | SLR |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Machine learning on thyroid disease: a review 15 | 2020 or later | ML algorithms differ in data. | ✓ |  |  |  |  |
| Application of Data Mining Techniques in Diagnosing Various Thyroid Ailments: A Review 16 | Not specified | data mining methods | ✓ |  |  |  |  |
| review of Deep Learning Approaches for Thyroid Cancer Diagnosis 17 | 2018 or late | Estimating diagnostic accuracy of deep learning | ✓ |  |  |  |  |
| Our study | 2013 to 2022 | thyroid disease diagnosis based on machine learning | ✓ | ✓ | ✓ | ✓ | ✓ |

# Method:

An SLR positions research questions before systematically searching for, selecting, and evaluating studies to see what information may be obtained from them18. This approach is chosen due to its reputation for providing a precise and reliable synthesis of scholarly content and is widely recognized across diverse research fields. The eligible studies items for meta-analyses, and systematic reviews or PRISMA, guidelines are followed while presenting the SLR.

## **Identification of the data:**

A thorough exploration was conducted using Scopus's integrated and WoS databases, which include all major publishers, including Emerald, Taylor & Springer, IEEE, and Willey. Many researchers consider the WoS and Scopus databases reliable for SLR due to the excellent quality of the indexing contents 19. The search covers 2013 to 2022 and includes all essential papers published during this time. Utilized terms such as "thyroid", "machine learning", "imbalance", and "deep learning" to find relevant publications. Boolean operators and various keywords are used to improve the search.

## **Screening initial data and determining eligibility:**

Searching for articles related to using ML techniques for diagnosing thyroid disease. The initial search using specific keywords on the Scopus database returned 2,182 articles and 486 articles on the WoS database. However, after applying the year limits of 2013-November-2022, the number of papers was narrowed down to 1159. Then 168 articles were selected for further screening based on document type, language, subject area, and keyword restrictions. The information for these 168 articles was exported as a CSV file in Excel, eliminating duplicates or leaving blanks.

Besides, reviewed the titles and abstracts of the remaining 168 unique papers. A standardized extraction form was used to extract the most relevant information. Research unrelated to machine learning but focused on thyroid disease was excluded. The researchers also decided not to include book chapters, ultrasound imaging, non-human studies, or reviews in their analysis. Figure 1 shows that all inclusion criteria were met by the 41 full-text publications that made it through the initial title and abstract screening.

# Observations and findings:

The following section will discuss the findings and insights gained by analyzing the metadata. These results are based on a meta-study of 168 papers, an analysis of their corresponding metadata, and a content analysis of 41 publications.

## **Metadata analysis:**

Metadata analysis helps understand scholarly literature by extracting information about the scholarly process's authors, articles, journals, and other elements 20. The metadata analysis was applied to 168 papers, and the papers were classified based on various factors, including a year of publication, publication type, publisher, country of origin, subject matter, funding source, and academic institution.

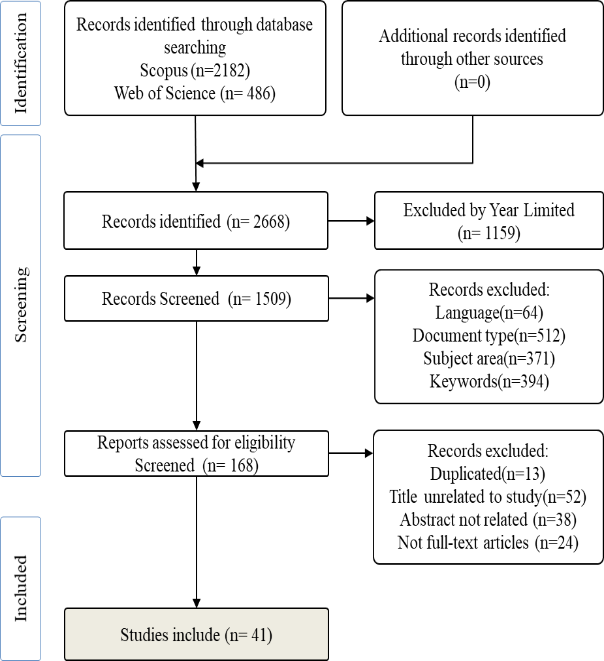


Figure PRISMA approach utilized in this research

## **Published by year:**

As shown in Figure 2, 168 papers were reviewed to see how many dealt with thyroid disease prediction using ML algorithms over the past decade. Publishing was expanding at a constant rate, and it is expected that this increase will be intensified significantly in the years 2020 and 2022. For example, in 2022, there were around 60 new papers published; in 2020, there were just 36 new papers published.

In addition, it has been evident throughout the period that the significance of the classification problem in the diagnosis of thyroid disease has received much attention. As a result, the number of scholarly works distributed to the public in 2022 is substantially more than in any previous year. On the other hand, one can observe the minuscule amount of papers published, particularly from 2013 to 2017, when there were merely a handful of papers. Consequently, increasing focus and concern are directed toward diagnosing thyroid disease, including classification issues and other data-driven concerns.

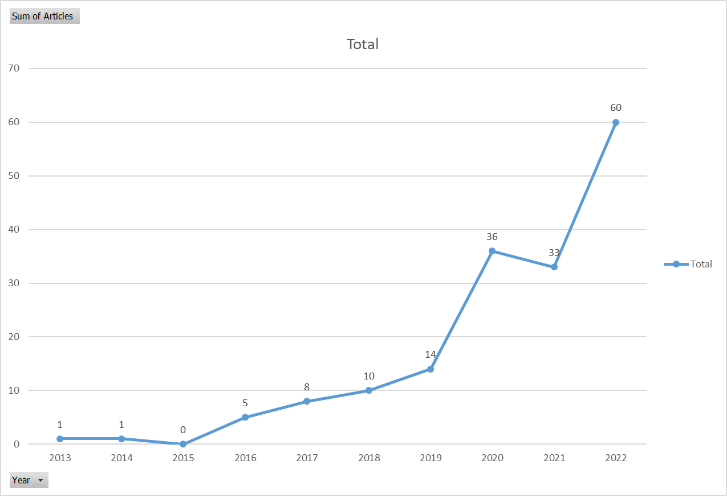


Figure Published papers on thyroid disease by using ML

**Most Relevant Authors:**

According to Figure 3, Fu C and Liu W. have penned the most pertinent papers of the five and are, therefore, the most impactful authors. Our team conducted a comparable data examination to monitor the authors' production over time. Indeed, findings revealed that Fu C and Liu W jointly produced four 2021 articles that received 10.5 citations.

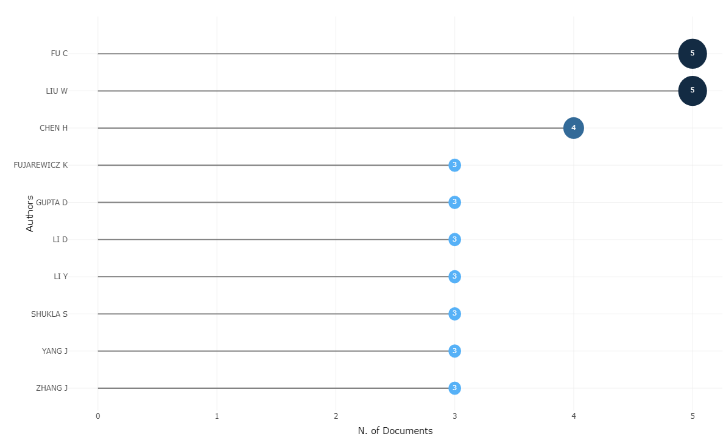
****

Figure Most relevant authors. The x-axis indicates the number of documents; the y-axis indicates

### **Most Relevant Sources:**

As shown in Figure 4, the most pertinent sources had ten documents: advances in intelligent systems and computers, expert systems with applications, a total of 6, and The Journal of The study material in networks and systems, a total of 5. Similarly.

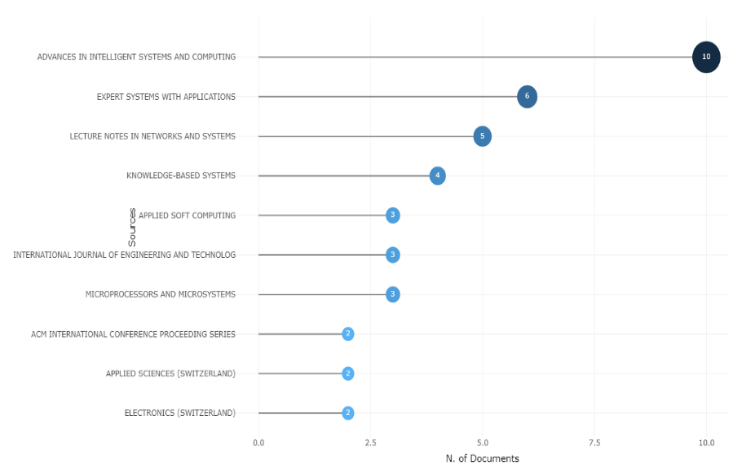
****

Figure Most relevant sources. The x-axis shows the number of documents; the y-axis indicates

### **Most frequently words used in the titles and keywords Table:**

The most common single, double, and triple keywords employed in the paper titles are shown in Table 2. The R-software program is used to identify the most popular keywords. Even though our main goal was to find and analyze articles that only focused on concepts like machine learning, deep learning, imbalance class, and thyroid disease, It came as a shock to discover that only "thyroid" and "disease" emerged as the top most frequently utilized terms in the keywords listed in Table 3. The writers used the terms "machine learning" 31 times in the keyword field, followed by the terms "thyroid disease" 24 times and "classification" 19 times. Articles often use the exact phrases which are listed below. However, intriguing outcomes were discovered when the examination was limited to the keywords utilized by the authors in the articles' keyword sections. A word cloud is a straightforward method for identifying the prevalent themes and key phrases in the referenced articles, allowing for the identification of the most general terms in a complex environment. Figure 5 displays word clouds generated by software, where larger and bolder text represents the terms most frequently used, and smaller and less bold text highlights the less commonly used phrases.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Unigrams | Frequency | Bigrams | Frequency | Trigrams | Frequency |
| thyroid | 86 | thyroid disease | 29 | thyroid disease prediction | 5 |
| learning | 43 | machine learning | 25 | Machine learning algorithms | 5 |
| classification | 42 | Deep  learning | 9 | support vector machine | 4 |
| disease | 36 | feature selection | 9 | deep neural network | 3 |
| machine | 32 | thyroid cancer | 9 | machine learning models | 3 |
| data | 26 | neural network | 8 | machine learning techniques | 3 |
| based | 24 | disease diagnosis | 6 | medical data classification | 3 |
| diagnosis | 23 | disease prediction | 5 | thyroid disease classification | 3 |
| cancer | 18 | learning algorithms | 5 | thyroid disease diagnosis | 3 |
| deep | 18 | support vector | 5 | artificial neural network | 2 |

Table most frequently utilized words in keyword sections

|  |  |
| --- | --- |
| Terms | Frequency |
| machine learning | 31 |
| thyroid disease | 24 |
| classification | 19 |
| thyroid | 18 |
| feature selection | 16 |
| deep learning | 15 |
| data mining | 14 |
| random forest | 10 |
| decision tree | 9 |
| thyroid cancer | 8 |

Table Terms repeated in articles

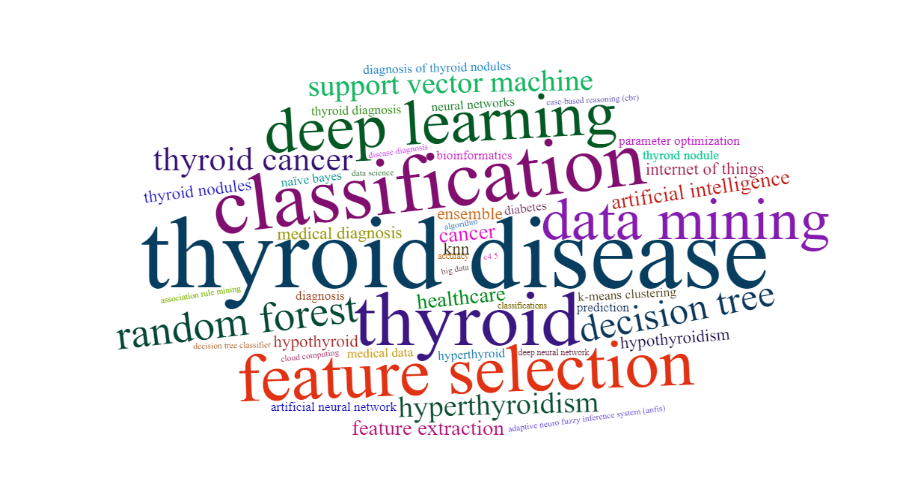


Figure Word cloud for most frequently used keywords in thyroid disease publications.

**Trending Topics:**

Our research allowed us to identify the prevailing themes concealed within the dataset's titles and abstracts. There were 68 occurrences of the word "thyroid" in the titles. Figure 6 displays that "classification" was the second most popular theme in the literature, appearing 42 times.

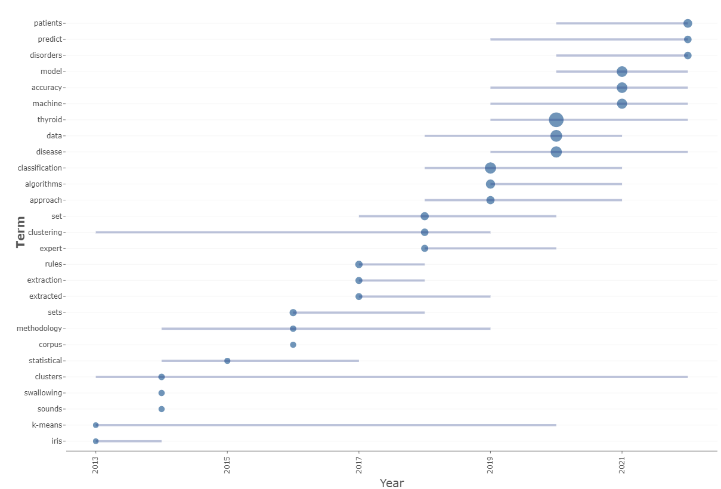


Figure Trending topics extracted from the topic of thyroid disease prediction

**Publication by institutions:**

Figure 7 shows the most cited articles sorted by the institution where their authors work. Regarding the number of publications related to thyroid, the Figure indicates that the leader is the Hefei University of Technology in China. In total, 14 articles came out of the institutes. Then the second highest number of publications came from China's Jilin University, with 13.

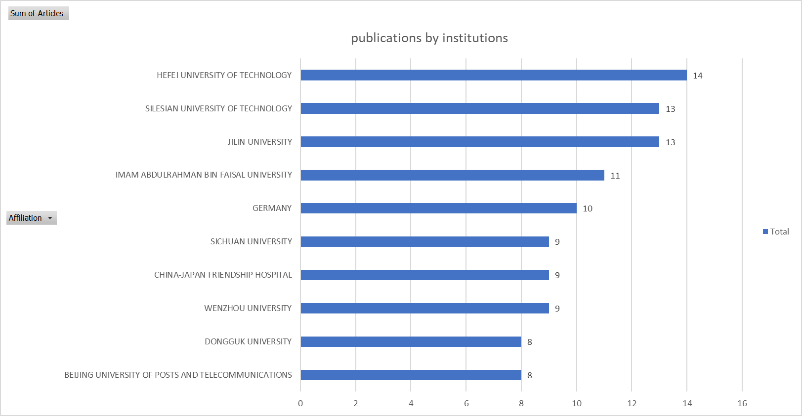


Figure Top ten institutions based on the number of publications

## **Insights of TDDBML:**

In this section, an in-depth examination of 41 research articles will be conducted, covering topics such as unbalanced data, thyroid disease, and machine learning. This review aims to provide insight into the concepts, methods, and potential future applications relevant to both theorists and practitioners.

## **Thyroid disease kinds:**

As machine learning-based methods improved, scientists and doctors began using data-driven methods to determine if a patient had a thyroid issue from a blood sample. Patients, however, often have to wait until their symptoms have worsened before they see a doctor because of the difficulties involved in undertaking the numerous routine tests. However, ML-based methods enable early-stage diagnosis, which the subject himself be able to routinely perform using inexpensive and compact sensors10. Hypothyroidism, euthyroidism, goiter, thyroiditis, thyroid cancer, thyroid hormone resistance, and hyperthyroidism are seven diagnostic categories for thyroid disease. There are two main types of hypothyroidism and hyperthyroidism. At least 15 of the papers out of the 41 chosen ones took into account two types because of their fatal consequences. Both types affect metabolism function, and conditions are severe and need medical attention21.

Indeed, thyroid disease, particularly in its terminal stage, is related to an increased risk of cardiovascular illness, elevated blood pressure, higher cholesterol levels, and mental depression22. In order to effectively treat patients with thyroid disease, it is crucial to diagnose the condition early. For instance, Ahmed et al. achieved a 98.2% accuracy rate when training a deep neural network to differentiate between hypothyroid and hyperthyroid states23. Pal et al. compared the three machine learning models for predicting thyroid disease, including KNN, DT, and multilayer perceptron (MLP), and found that it achieves the highest accuracy of 94.23% 24. The UCI thyroid disease open repository dataset was used in the study. On the other hand, Aljameel used an EANN-based approach to distinguish between thyroid cancer and non-cancer raw data using real-world data with 99% accuracy25.

Figure 8 depicts the most commonly reported disease associated with the thyroid from cited studies. Compared to other key-related diseases, "thyroid disease" has the largest cluster in the Figure; the two most common types of thyroid disease are also displayed, hyperthyroidism and hypothyroidism. In addition, several keywords have been repeated, which indicate that it is a technique used to predict a thyroid disease in the early stage, such as random forest, k-NN, and ANN.

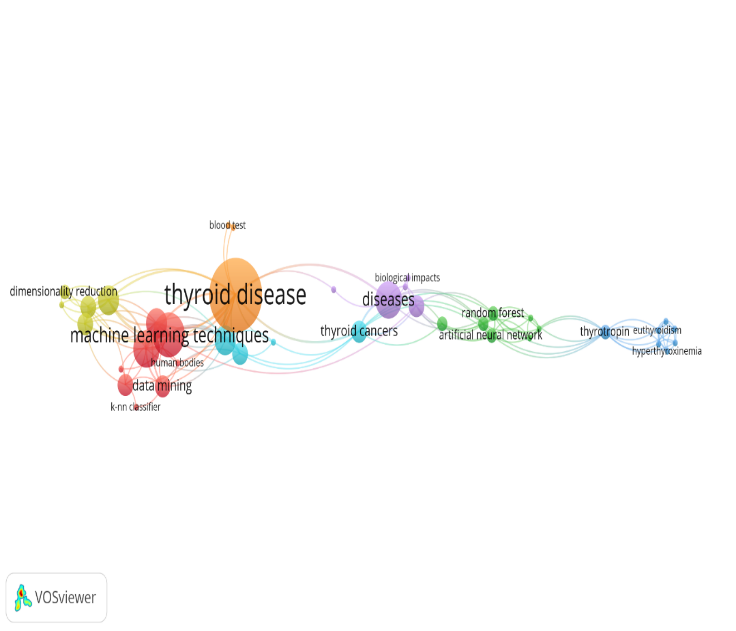


Figure Illustration of most-reported thyroid disease (developed by VOSviewer software).

**Machine learning algorithms:**

Table 4 illustrates that support vector machine (SVM) algorithms have received more attention from researchers and practitioners than any other ML type in designing PTDBML models. At least 12 of the 41 studies that attempted to develop a model to diagnose thyroid disease used an SVM-based approach that used the standard technique in healthcare system prediction26. For instance, Płuciennik et al. have developed a model for thyroid cancer diagnostics, which achieved approximately 95% accuracy27. Vairale et al. compared SVM to Logistic Regression (LR), K-NN, and NN for identifying people with the hypothyroid disease on the actual case dataset. SVM showed the best performance among all algorithms, producing an accuracy of 99%28.

On the Other side, the RF classifier is the following algorithm to enhance the thyroid disease prediction model, which was nine studies conducted to develop a model for thyroid disease prediction. Alghamdi has designed an efficient predictive model to find thyroid cancer in the  Prostate, Lung, Colorectal, and Ovarian (PLCO) dataset, defined as 155000 examples29. They used seven models the Logistic Regression model (LR), KNN, Ada boost classifier (AdaB), SVM, DT, Gaussian Naïve Bayes (GNB), RF, and Gradient Boosting classifier (GB); the RF has vital accuracy of 100%.

There are evident that, as time has progressed, a growing number of TDDBML model development efforts have focused on DL algorithms rather than classic ML. Only 8 out of 41 studies focused on using DL to create a model for TDDBML, indicating that more research is needed. In order to classify individuals into normal, hyperthyroid, and hypothyroid categories, Guleria et al. used a thyroid cancer prediction system based on MLP. According to preliminary computational results, the proposed model identifies thyroid issues with an accuracy of 99.8% 30. M Asif et al. proposed that a multilayer Perceptron (MLPC) was the most effective algorithm, achieving an accuracy of 99.70% 31. In addition, Zhou et al. used ten ML algorithms through thyroid surgery to demonstrate a corresponding model. A CNN model can use Auc and accuracy to identify patients at an early stage of thyroid disease. Their main finding is that using data from 500 actual patients, the model can accurately identify those with thyroid disease 90% of the time and with 83% Auc. Other ML-based algorithms used by researchers to create the TDDBML model include KNN 9,11,31, Hoeffding 33, XGBoost 22, and Adaboost and Bagging 34.

### **Imbalance challenges:**

One of the initial focuses was to track down previous research publications on the field of thyroid disease that included analyses of imbalanced data. However, it became evident as one read through the articles that the vast majority of research either adopted data from other open sources or their studies used actual data and that in both situations, the datasets were unbalanced. As a result, the quality assessment revealed that eight articles relied on experimental results from the unbalanced dataset. Recent studies have addressed the issue of imbalanced data's effect on model performances, which most of the studies ignored.

The imbalance problems are dealt with in various ways depending on the author. For instance, Zhou et al. assessed the model performance on unbalanced data classification by computing its f1 score, ROC-AUC curves, and accuracy rate 35. N. Alghamdi has worked on the PLCO dataset, which shows patients that more classes have not been diagnosed with thyroid cancer, and fewer classes are diagnosed with thyroid cancer, and they relied on an under-sampling technique to handle imbalanced classes29. S. Aljameel et al. worked on a dataset that had an imbalance (much more thyroid cancer cases than nonthyroid cancer cases); thus, they utilized the SMOTEENN technique to avoid biasing the models toward one of the outcomes25. While 9,36,37 rely on SMOTE to handle imbalanced data issues to prevent bias in the performance measures. In cases where SMOTE is used to equalize the data, the overall model accuracy increases. Hayashi et al. have suggested a model like continuous Re-RX extract informative principles from the thyroid dataset with the correct values of subdivision rate for both the majority and minority classes 38.

Some researchers adopt DL-based solutions to replace all other algorithm-level methods. For instance, Selwal & Raoof developed a more accurate thyroid disease prediction system was used the MLP machine learning model and tested it on random samples that included hyperthyroid, hypothyroid, and healthy subjects30. On the other hand, after choosing variables for thyroid illness prediction, several studies employ the Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and CNNLSTM. The authors demonstrate that AUC of 72% may be achieved using their proposed model. Thyroid disease-related datasets are notoriously unbalanced, and few publications have investigated methods to address this issue outside of classification and fabricated models10,15,39. Many ways are available to handle imbalanced classes; however, few studies have been mentioned in this overview that impact model performance the most.

# Discussions:

A thorough examination of 41 studies was done to understand the current practices and techniques used for identifying thyroid disorders when working with an unbalanced dataset. The comprehensive analysis evaluated the following factors: thyroid disease type, applications, machine learning (ML) algorithms, and imbalance solutions.

Overall, hypothyroidism and hyperthyroidism have received excellent attention in TDDBML. At the same time, other investigations looked into euthyroid 9, thyroid surgery35, and so on. Sick-euthyroid syndrome is one of the common problems in people suffering from malnutrition, trauma, surgery, or severe acute or chronic disease. Consequently, researchers and practitioners may focus on all types of thyroid disease rather than hypothyroidism and hyperthyroidism. Most ML-based models are designed to detect thyroid disease patients, emphasizing classification. Because of their availability and the issues connected with data imbalance concerns, most researchers heavily investigated popular datasets: UC Irvine Machine Learning Repository. However, a few research took into account real-world data28,35,40 and large datasets30,39. A large amount of data helps the healthcare industry create more effective disease detection and decision-support systems26. The performance variation of the model is detected in the study findings supplied as public source data and actual data. However, it cannot be denied that the performance of the models will be more accurate when the experiment is conducted using actual data. In order to evaluate the effectiveness of ML-based models, it is required to use actual data rather than public repository data.

The instability of the model is one of the primary factors contributing to the restricted capabilities of the Clinical Decision Support Systems (CDSS) system. Since clinical systems cannot function correctly using only old patient data, the CDSS model must be continuously refined and updated, considering new information. Situations where it is necessary to collect data in real-time and train an ML model, such as the operating room during an emergency or a blood test conducted with the new devices, are likely to provide significant difficulties.

Several machine learning (ML) methods, including SVM, RF, DT, KNN, ANN, MLP, and NB, are employed to construct the TDDBML model. However, the RF-based model attracts the most excellent attention from researchers since it provides the most accurate classification. Thousands of variables can be accommodated by random forest. Using data level, the problems of data ratio imbalance have been mainly addressed. SMOTE remains the most popular data-level solution, as seen by citations from only a few years ago. Unbalanced class solutions, which are more common in recent publications, have also contributed to the popularity of the DL-based method among academics. One of the problems with the DL-based method is that it does not explain how the model arrives at its conclusions and struggles with overfitting, where they memorize the training data too well and do not perform well on new data 41.

Machine learning models often perform better intra-patient than inter-patient (inter-patient). Different data or patient characteristics can cause this. If a machine learning model is trained and evaluated on a dataset of individuals with one type of thyroid disease, it may not perform as well on another dataset. This could be due to patient-specific data, such as symptoms or blood test findings. To increase a machine learning model's performance on inter-patient data, a larger diverse dataset of patients may be needed to train the model to generalize to a broader range of patient populations. Choosing a different machine-learning method may be essential.

Most traditional classification methods try to find an ideal classifier that maximizes classification accuracy while keeping the misclassification cost constant, which can be problematic when dealing with imbalanced classes42, considering the potential that the cost of misclassification may vary based on the probability distribution of the sample. In addition, most of the reported research included computationally costly techniques, including noising, thyroid segmentation, feature extractions, and classifications43. Implementing such a model in the actual world would be difficult and might be a fascinating area for future research. This is particularly relevant when the consequences of misclassification are severe, such as in medical diagnostics.

In addition, there are challenges related to machine learning models' accuracy and reliability. Machine learning models are only as good as the data they are trained on, and if the data is of poor quality or biased, the model's predictions may not be accurate 44. It is essential to carefully evaluate the performance of machine learning models using appropriate evaluation metrics and consider the models' limitations when making predictions. Clinical diagnosis systems based on machine learning raise security problems for making diagnoses.

The model's accuracy may vary depending on factors such as the location's geology, the size of the data set used, and the different types of thyroid disease being modeled. Some ML models, like those used to identify hypothyroidism, may not apply to other conditions. Since the training data for each disease is likely to come from a different source, developing an ML-based clinical detection method may require several models. This can make developing a framework to deliver accurate and real-time disease determinations complex and challenging. The stability of the model through variable parameters or model updates based on user experience is also crucial to a safe diagnosis. (Table 4) collects the results of the 41 studies cited in the literature to provide light on the ML-based prediction of thyroid disease.

Table Literature Emphasized On TDDBML

| Author | Title | Algorithms | Imbalance | Evaluation | Dataset |
| --- | --- | --- | --- | --- | --- |
| 27 | Data Integration–Possibilities of Molecular and Clinical Data Fusion on the Example of Thyroid Cancer Diagnostics | SVM classifier (linear kernel) | NA | p value below 0.05 | 200 real case |
| 35 | Predicting difficult airway intubation in thyroid surgery using multiple machine learning and deep learning algorithms | LR, RF, GB, XGB,LGBM,MLPC,GNB,CNN,LSTM, CNNLSTM | ROCNA AUC curves | 92,91,91,91,91,90, 89,90, 89, 90 | 500 real case |
| 40 | Early diagnosis of thyroid cancer diseases using computational intelligence techniques: A case study of a Saudi Arabian dataset | RF, SVM, ANN, and NB | NA | 90, 84, 88, 81 | 218 real case |
| 29 | Evaluation of classification models for predicting mortality rate using thyroid cancer data | LR, KN, SVC, GNB,DT, AdaB ,RF and GB | Under sampling | 98, 98, 99,  99,99, 99, 1, 1 | 155000 |
| 25 | A Proactive Explainable Artificial Neural Network Model for the Early Diagnosis of Thyroid Cancer | explainable artificial neural network (EANN) | SMOTEENN | 0.98 | 724 real case |
| 9 | Application of machine learning algorithms to predict the thyroid disease risk: an experimental comparative study | ANN, CatBoost, XGB, RF, LGBM, DT, SVC, KNN ,GNB | SMOTE | 95, 95, 95, 94,94, 94, 91, 89, 86 | 3,162  UCI |
| 22 | Thyroid Disease Prediction Using XGBoost Algorithms | LR, DT, KNN, and XGB | NA | 81, 87, 96, 98 | 215 UCI |
| 45 | of thyroid disorder Effective voting ensemble of homogenous ensembling with multiple attribute selection approaches for improved identification | DT, GB,LR, RF | NA | 1 | 309 UCI |
| 38 | Use of the recursive rule extraction algorithm with continuous attributes to improve diagnostic accuracy in thyroid disease | Re RX | BRACID | 96.70 | 7200 UCI |
| 46 | Butterfly Optimized Feature Selection with Fuzzy CNA Means Classifier for Thyroid Prediction | fuzzy CNA means algorithm (FCM) | NA | 0.943 | 4152 UCI |
| 47 | Expanded and Filtered Features Based ELM Model for Thyroid Disease Classification | ELM | NA | 99.68 | 12944 UCI |
| 39 | Developing an Explainable Machine Learning Based Thyroid Disease Prediction Model | LR | NA | 91 | 215 UCI |
| 32 | Increasing the Prediction Accuracy for Thyroid Disease: A Step Towards Better Health for Society | KNN  NN | NA | 94, 98 | 3152 UCI |
| 30 | A Multi- layer perceptron-based intelligent thyroid disease prediction system | MLP | NA | 99 | 120 UCI |
| 11 | Early prediction of hypothyroidism and multiclass classification using predictive machine learning and deep learning | DT, ANN | NA | 99, 99 | 3772  UCI |
| 10 | Constituent depletion and divination of hypothyroid prevalance using machine learning classification | RF, DT, NB, KNN, LR | NA | 99 | 3164 UCI |
| 48 | Constructing a system for analysis of machine learning techniques for early detection of thyroid | ZeroR J48  Naïve bayes OneR | NA | 60, 68, 41, 64 | 1,000  UCI |
| 33 | Decision tree ensemble techniques to predict thyroid disease | J48, RT, Hoeffding | NA | 99, 97, 92 | 499 UCI |
| 43 | Prediction of Thyroid isease(Hypothyroid) in Early Stage Using Feature Selection and Classification Techniques | SVM, DT, RF, LR, NB | NA | 99, 99, 99, 96 | 519 UCI |
| 36 | Thy- Sys: A Preliminary Thyroid Wellness Assessment Through Machine Learning Using Pathological Factors | SVM, KNN, DT, SVMNA KNN | SMOTE | 99.5 | 1464  UCI |
| 49 | Predictive Analysis for Thyroid Diseases Diagnosis Using Machine Learning | KNN, NB, DT | NA | 92, 95, 99 | 1464 UCI |
| 31 | Computer-aided diagnosis of thyroid disease using machine learning algorithms | KNN, SVM, AdB, XGB, GPC, GBC,MLPC | NA | 93, 96, 97, 96, 95, 98, 99 | 3164  UCI |
| 50 | A Machine Learning Approach to Predict Thyroid Disease at Early Stages of Diagnosis | DT, NB | NA | 95 | 3000  UCI |
| 51 | Prediction of thyroid disorders using advanced machine learning techniques | NB, SVM, RF | NA | 74, 92, 78 | 7200  UCI |
| 28 | Classification of Hypothyroid Disorder using Optimized SVM Method | KNN, SVM, LR, NN | NA | 97, 99, 95, 94 | 574 real case |
| 52 | A Study on Label TSH, T3, T4U, TT4, FTI in Hyperthyroidism and Hypothyroidism using Machine Learning Techniques | RF, SVM, KNN | NA | 98, 97, 95 | 7200  UCI |
| 53 | Feature selection algorithms to improve thyroid disease diagnosis | MLP BPNN SVM ELM | NA | 94, 95, 97, 98 | 215 UCI |
| 24 | Enhanced Prediction of Thyroid Disease Using Machine Learning Method | KNN, DT, MLP | NA | 91, 94, 96 | 3163  UCI |
| 37 | Accuracy Assessment of Machine Learning Algorithm(s) in Thyroid Dysfunction Diagnosis | J48, MLP, NB, RF, SVM | SMOTE | 99, 98, 98, 99, 98 | 4975  UCI |
| 34 | Efficient Thyroid Disease Prediction using Features Selection and Meta-Classifiers | Ada Boosting Bagging | NA | 93, 99 | 774 |

# Conclusions

This study seeks to discover the most recent ML-based and data-driven developments and strategies in diagnosing thyroid disease using imbalanced data. To develop ML-based systems for predicting thyroid disease in the real world, it is essential to enhance the ML-based experiments to include real-data patients and interpretable machine learning to explain the final prediction adequately. A comprehensive review of 41 papers suggests that more research is needed to prove reliable performance in healthcare settings. Although Deep Learning has come to dominate the area, SMOTE is still widely used as an Over-Sampling technique for handling unbalanced data by many academics and practitioners. Many researchers have noticed the development of an RF-based model for predicting thyroid disease since it is easier to train and can handle many features. Another big attraction is that they resist overfitting, making them useful in various machine-learning applications. The limits of ML that are discussed in the discussion sections may guide the direction of future research. Regardless, it is expected that ML-based thyroid disease detection utilizing imbalanced data and innovative techniques will uncover numerous undiscovered opportunities in the future.

**Author's declaration:**

* Conflicts of Interest: None.
* We hereby confirm that all the Figures and Tables in the manuscript are mine.
* Ethical Clearance: The project was approved by the local ethical committee at University Technology Malaysia.

**Authors' contributions statement:**

Dh. S. S.: He conceived and designed the paper, and the paper was my idea. He made the acquisition and, analysis, and interpretation of the data. M. Sh. O: He did Conception, design of the work and critical revision of the article. He did approval of the version to be published.

**References:**

1. Clinic M. Hypothyroidism (underactive thyroid). Published online 2021. https://www.mayoclinic.org/diseases-conditions/hypothyroidism/symptoms-causes/syc-20350284

2. Thabit MA, Abdullah GH, Al-Rawi KF. Polymorphism study of MTHFR 677C?T and its correlation with oxidative stress and their influence on female infertility in Erbil - Iraq. *Baghdad Sci J*. 2017;14(3):611-618. doi:10.21123/bsj.2017.14.3.0611

3. World Health Organization (WHO). pepole death by Thyroid diseases. Published online 2020.

4. Hamamurad QH, Jusoh NM, Ujang U. Modern City Issues, Management and the Critical Role of Information and Communication Technology. *Int J Adv Comput Sci Appl*. 2022;13(4):368-373. doi:10.14569/IJACSA.2022.0130443

5. American Cancer Society. Thyroid Cancer. Published 2021. https://www.cancer.org/cancer/thyroid-cancer.html

6. Obschonka M, Audretsch DB. Artificial intelligence and big data in entrepreneurship: a new era has begun. *Small Bus Econ*. 2020;55(3):529-539. doi:10.1007/s11187-019-00202-4

7. Garbuio M, Lin N. Artificial intelligence as a growth engine for health care startups: Emerging business models. *Calif Manage Rev*. 2019;61(2):59-83. doi:10.1177/0008125618811931

8. Alfifi M, Alrahhal MS, Bataineh S, Mezher M. Enhanced artificial intelligence system for diagnosing and predicting breast cancer using deep learning. *Int J Adv Comput Sci Appl*. 2020;11(7):498-513. doi:10.14569/IJACSA.2020.0110763

9. Islam SS, Haque MS, Miah MSU, Sarwar T Bin, Nugraha R. Application of machine learning algorithms to predict the thyroid disease risk: an experimental comparative study. *PeerJ Comput Sci*. 2022;8:1-35. doi:10.7717/PEERJ-CS.898

10. Shyamala Devi M, Shil A, Katyayan P, Surana T. Constituent depletion and divination of hypothyroid prevalance using machine learning classification. *Int J Innov Technol Explor Eng*. 2019;8(12):1607-1612. doi:10.35940/ijitee.L3150.1081219

11. Guleria K, Sharma S, Kumar S, Tiwari S. Early prediction of hypothyroidism and multiclass classification using predictive machine learning and deep learning. *Meas Sensors*. 2022;24(September):100482. doi:10.1016/j.measen.2022.100482

12. Firestone AJ, Settleman J. A three-drug combination to treat BRAF-mutant cancers. *Nat Med*. 2017;23(8):913-914.

13. Ahsan MM, Ahad MT, Soma FA, et al. Detecting SARS-CoV-2 from chest X-Ray using artificial intelligence. *Ieee Access*. 2021;9:35501-35513.

14. Feng W, Huang W, Ren J. Class imbalance ensemble learning based on the margin theory. *Appl Sci*. 2018;8(5):815.

15. Lee KS, Park H. Machine learning on thyroid disease: a review. *Front Biosci - Landmark*. 2022;27(3). doi:10.31083/j.fbl2703101

16. Mendoza AM, Hernandez RM. Application of Data Mining Techniques in Diagnosing Various Thyroid Ailments: A Review. *Proc 2021 13th Int Conf Inf Commun Technol Syst ICTS 2021*. Published online 2021:207-212. doi:10.1109/ICTS52701.2021.9608400

17. Anari S, Tataei Sarshar N, Mahjoori N, Dorosti S, Rezaie A. Review of Deep Learning Approaches for Thyroid Cancer Diagnosis. *Math Probl Eng*. 2022;2022. doi:10.1155/2022/5052435

18. Okoli C, Schabram K. Working Papers on Information Systems A Guide to Conducting a Systematic Literature Review of Information Systems Research. *Work Pap Inf Syst*. 2010;10(2010). doi:10.2139/ssrn.1954824

19. Fahimnia B, Sarkis J, Davarzani H. Green supply chain management: A review and bibliometric analysis. *Int J Prod Econ*. 2015;162:101-114. doi:10.1016/j.ijpe.2015.01.003

20. Borgman CL. Big Data, Little Data, No Data: Scholarship in the Networked World—Who is in Charge of Data Quality. Published online 2015.

21. Journal BS. Effect of Hyper and Hypothyroidism on Lipid Profile and Liver Function of Male Rats. *Baghdad Sci J*. 2011;8(4):926-933. doi:10.21123/bsj.8.4.926-933

22. Sankar S, Potti A, Naga Chandrika G, Ramasubbareddy S. Thyroid Disease Prediction Using XGBoost Algorithms. *J Mob Multimed*. 2022;18(3):917-934. doi:10.13052/jmm1550-4646.18322

23. Ahmed I, Mohiuddin R, Muqeet MA, Kumar JA, Thaniserikaran A. Thyroid Cancer Detection using Deep Neural Network. *Proc - Int Conf Appl Artif Intell Comput ICAAIC 2022*. 2022;(Icaaic):166-169. doi:10.1109/ICAAIC53929.2022.9792854

24. Pal M, Parija S, Panda G. Enhanced Prediction of Thyroid Disease Using Machine Learning Method. *Proc IEEE VLSI DCS 2022 3rd IEEE Conf VLSI Device, Circuit Syst*. 2022;2022(February):199-204. doi:10.1109/VLSIDCS53788.2022.9811472

25. Aljameel SS. A Proactive Explainable Artificial Neural Network Model for the Early Diagnosis of Thyroid Cancer. *Computation*. 2022;10(10). doi:10.3390/computation10100183

26. Kamra V, Kumar P, Mohammadian M. Diagnosis support system for general diseases by implementing a novel machine learning based classifier. *Int J Comput Digit Syst*. 2021;10(1):737-746. doi:10.12785/ijcds/100168

27. Płuciennik A, Płaczek A, Wilk A, Student S, Oczko-Wojciechowska M, Fujarewicz K. Data Integration–Possibilities of Molecular and Clinical Data Fusion on the Example of Thyroid Cancer Diagnostics. *Int J Mol Sci*. 2022;23(19). doi:10.3390/ijms231911880

28. Vairale VS, Shukla S. Classification of Hypothyroid Disorder using Optimized SVM Method. *Proc 2nd Int Conf Smart Syst Inven Technol ICSSIT 2019*. 2019;(Icssit):258-263. doi:10.1109/ICSSIT46314.2019.8987767

29. Alghamdi NS. Evaluation of classification models for predicting mortality rate using thyroid cancer data. *J Comput Sci*. 2019;15(1):131-142. doi:10.3844/jcssp.2019.131.142

30. Selwal A, Raoof I. A Multi-layer perceptron based intelligent thyroid disease prediction system. *Indones J Electr Eng Comput Sci*. 2020;17(1):524-532. doi:10.11591/ijeecs.v17.i1.pp524-532

31. Asif MAAR, Nishat MM, Faisal F, et al. Computer aided diagnosis of thyroid disease using machine learning algorithms. *Proc 2020 11th Int Conf Electr Comput Eng ICECE 2020*. 2020;4:222-225. doi:10.1109/ICECE51571.2020.9393054

32. Jha R, Bhattacharjee V, Mustafi A. Increasing the Prediction Accuracy for Thyroid Disease: A Step Towards Better Health for Society. *Wirel Pers Commun*. 2022;122(2):1921-1938. doi:10.1007/s11277-021-08974-3

33. Yadav DC, Pal S. Decision tree ensemble techniques to predict thyroid disease. *Int J Recent Technol Eng*. 2019;8(3):8242-8246. doi:10.35940/ijrte.C6727.098319

34. Priyadharsini D, Sasikala S. Efficient Thyroid Disease Prediction using Features Selection and Meta-Classifiers. *Proc - 6th Int Conf Comput Methodol Commun ICCMC 2022*. 2022;(Iccmc):1236-1243. doi:10.1109/ICCMC53470.2022.9753986

35. Zhou CM, Wang Y, Xue Q, Yang JJ, Zhu Y. Predicting difficult airway intubation in thyroid surgery using multiple machine learning and deep learning algorithms. *Front Public Heal*. 2022;10. doi:10.3389/fpubh.2022.937471

36. Francisco IR, Ferolin MBJ, Pena CF, Ferolin RJ. Thy-Sys: A Preliminary Thyroid Wellness Assessment Through Machine Learning Using Pathological Factors. *Proc - 2021 1st Int Conf Inf Comput Res iCORE 2021*. Published online 2021:44-49. doi:10.1109/iCORE54267.2021.00027

37. Danjuma KJ, Maksha Wajiga G, Garba EJ, Sandra Ahmadu A, Longe OB. Accuracy Assessment of Machine Learning Algorithm(s) in Thyroid Dysfunction Diagnosis. *Proc 2022 IEEE Niger 4th Int Conf Disruptive Technol Sustain Dev NIGERCON 2022*. Published online 2022. doi:10.1109/NIGERCON54645.2022.9803113

38. Hayashi Y, Nakano S, Fujisawa S. Use of the recursive-rule extraction algorithm with continuous attributes to improve diagnostic accuracy in thyroid disease. *Informatics Med Unlocked*. 2015;1(2015):1-8. doi:10.1016/j.imu.2015.12.003

39. Arjaria SK, Rathore AS, Chaubey G. Developing an Explainable Machine Learning-Based Thyroid Disease Prediction Model. *Int J Bus Anal*. 2022;9(3):1-18. doi:10.4018/IJBAN.292058

40. Olatunji SO, Alotaibi S, Almutairi E, et al. Early diagnosis of thyroid cancer diseases using computational intelligence techniques: A case study of a Saudi Arabian dataset. *Comput Biol Med*. 2021;131(February):104267. doi:10.1016/j.compbiomed.2021.104267

41. Francois C. Deep Learning with Python, Manning Publications. Published online 2017.

42. Gan D, Shen J, An B, Xu M, Liu N. Integrating TANBN with cost sensitive classification algorithm for imbalanced data in medical diagnosis. *Comput Ind Eng*. 2020;140(June 2019):106266. doi:10.1016/j.cie.2019.106266

43. Riajuliislam M, Rahim KZ, Mahmud A. Prediction of Thyroid Disease(Hypothyroid) in Early Stage Using Feature Selection and Classification Techniques. *2021 Int Conf Inf Commun Technol Sustain Dev ICICT4SD 2021 - Proc*. Published online 2021:60-64. doi:10.1109/ICICT4SD50815.2021.9397052

44. Kitchenham B, Charters S. Guidelines for performing systematic literature reviews in software engineering. Published online 2007.

45. Akhtar T, Gilani SO, Mushtaq Z, et al. Effective voting ensemble of homogenous ensembling with multiple attribute-selection approaches for improved identification of thyroid disorder. *Electron*. 2021;10(23). doi:10.3390/electronics10233026

46. J. K. Jagadeesh Kumar S, Parthasarathi P, Masud M, F. Al-Amri J, Abouhawwash M. Butterfly Optimized Feature Selection with Fuzzy C-Means Classifier for Thyroid Prediction. *Intell Autom Soft Comput*. 2023;35(3):2909-2924. doi:10.32604/iasc.2023.030335

47. Juneja K. *Expanded and Filtered Features Based ELM Model for Thyroid Disease Classification*. Springer US; 2022. doi:10.1007/s11277-022-09823-7

48. Rasheeduddin S, Rajasekhar Rao K. Constructing a system for analysis of machine learning techniques for early detection of thyroid. *Int J Eng Adv Technol*. 2019;8(6 Special Issue 3):1978-1981. doi:10.35940/ijeat.F1385.0986S319

49. Peya ZJ, Chumki MKN, Zaman KM. Predictive Analysis for Thyroid Diseases Diagnosis Using Machine Learning. *2021 Int Conf Sci Contemp Technol ICSCT 2021*. Published online 2021:4-9. doi:10.1109/ICSCT53883.2021.9642544

50. Rao AR, Renuka BS. A Machine Learning Approach to Predict Thyroid Disease at Early Stages of Diagnosis. *2020 IEEE Int Conf Innov Technol INOCON 2020*. Published online 2020:2020-2023. doi:10.1109/INOCON50539.2020.9298252

51. Duggal P, Shukla S. Prediction of thyroid disorders using advanced machine learning techniques. *Proc Conflu 2020 - 10th Int Conf Cloud Comput Data Sci Eng*. Published online 2020:670-675. doi:10.1109/Confluence47617.2020.9058102

52. Shahid AH, Singh MP, Raj RK, Suman R, Jawaid D, Alam M. A Study on Label TSH, T3, T4U, TT4, FTI in Hyperthyroidism and Hypothyroidism using Machine Learning Techniques. *Proc 4th Int Conf Commun Electron Syst ICCES 2019*. 2019;(Icces):930-933. doi:10.1109/ICCES45898.2019.9002284

53. Pavya K, Srinivasan B. Feature selection algorithms to improve thyroid disease diagnosis. *IEEE Int Conf Innov Green Energy Healthc Technol - 2017, IGEHT 2017*. Published online 2017:1-5. doi:10.1109/IGEHT.2017.8094070