



Article

Unveiling Thyroid Disease Associations: An Exceptionality-Based Data Mining Technique

Xinyu Zhang ^{1,*}, Vincent C. S. Lee ¹ and James C. Lee ^{2,3}

¹ Department of Data Science and AI, Faculty of IT, Monash University, Clayton, Melbourne, VIC 3800, Australia; vincent.cs.lee@monash.edu

² Monash University Endocrine Surgery Unit, Alfred Hospital, Melbourne, VIC 3004, Australia; james.lee@monash.edu

³ Department of Surgery, Central Clinical School, Monash University, Melbourne, VIC 3004, Australia

* Correspondence: xinyu.zhang@monash.edu; Tel.: +61-452295059

Abstract: Background: The prevalence of thyroid disease has seen a rapid increase in recent times, primarily attributed to the fast pace of lifestyles that often result in poor dietary choices, work-life imbalances, social stress, genetic mutations, and improved diagnostic capabilities. However, the precise contribution of these factors to thyroid disease remains a subject of controversy. Consequently, there is a pressing need to gain a comprehensive understanding of the related associations in order to potentially mitigate the associated morbidity and mortality rates. Methods: This study employed association rule mining techniques to reveal hidden correlations among complex and diverse epidemiological connections pertaining to thyroid disease associations. We proposed a framework which incorporates text mining and association rule mining algorithms with exceptionality measurement to simultaneously identify common and exception risk factors correlated with the disease through real-life digital health records. Two distinctive datasets were analyzed through two algorithms, and mutual factors were retained for interpretation. Results: The results confirmed that age, gender, and history of thyroid disease are risk factors positively related to subsequent thyroid cancer. Furthermore, it was observed that the absence of underlying chronic disease conditions, such as diabetes, hypertension, or obesity, are associated with reduced likelihood of being diagnosed with thyroid cancer. Conclusions: Collectively, the proposed framework demonstrates its sound feasibility and should be further recommended for different disease in-depth knowledge discovery.

Keywords: thyroid disease associations; association rule mining; text mining; common and exception rules; exceptionality measurement



Citation: Zhang, X.; Lee, V.C.S.; Lee, J.C. Unveiling Thyroid Disease Associations: An Exceptionality-Based Data Mining Technique. *Endocrines* **2023**, *4*, 558–572. <https://doi.org/10.3390/endocrines4030040>

Academic Editor: Alessandro Delitala

Received: 23 June 2023

Revised: 19 July 2023

Accepted: 26 July 2023

Published: 28 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The thyroid is one of the largest endocrine glands in the human body responsible for regulating the metabolism of cells [1]. Improper functioning of the gland might lead to functional disorders, such as hypothyroidism, hyperthyroidism, thyroiditis, or neoplastic thyroid diseases, like multi-nodular goiter, adenoma, or malignant tumors [2]. Thyroid disease incidence is continuously rising worldwide, and thyroid cancer is even the most rapidly increased malignancy in the past few decades [3]. Despite the increasing diagnostic instances, the cause and associations around the disease are still under-researched. Uncovering factors associated with the increased diagnosis of thyroid diseases may shed some light on potential contributors to this phenomenon.

Currently, a few studies adopted qualitative approaches to identify factors correlated with the disease. This kind of analysis takes long-established protocols to evaluate a potential factor at a time, which is time-consuming and generally ignores the potential correlations among diverse factors. Medical datasets are complex as many factors interweave one another, making the identification of thyroid disease associations even more challenging. Through such processes, the identified factors include radiation, depression,

obesity, hormonal factors, and gene heredity [4–8]. Nevertheless, the influencing effects of many of the factors are still under debate as they cannot be verified without quantitative determinations. Additionally, applying quantitative approaches for risk factors identification has been considerably ignored by existing studies, and this again aggravates the limited reliability and approbation degrees of the identified factors.

Association rule mining (ARM) techniques are responsible for revealing hidden patterns among diverse attributes given a database [9]. In the past few decades, ARM has shown effectiveness in identifying underlying correlations among comorbidity in the medical domain [9–11]. Medical records are high-dimensional and complex. Mining valuable knowledge manually from such heterogeneous data requires significant effort and time, which is challenging to accomplish. ARM, on the other hand, is considerably efficient for dealing with complex and sensitive records uptake by clinicians, making it decent for discovering unrevealed patterns from medical data.

The underlying patterns derived from ARM can be categorized into common rules (i.e., rules with high support and high confidence), reference rules (i.e., rules with low support and low confidence), and exception rules (i.e., rules with low support and high confidence) [12]. Common rules describe explicit information which interprets the regularity of objects with consequences. Reference rules are outliers that are less meaningful and are generally excluded. Exception rules outline the unexpectedness of associations and are often tied up with actionability. It should be noted that the rules with high support and low confidence were generally removed, as they were uncommon and considered uninteresting in clinical practice. Existing studies primarily focus on extracting common rules, whereas the extraction of exception rules has been considerably neglected. Nevertheless, exception rules are potentially more engaging and valuable than common rules [13] as they can provide information that reveals unusual and contradictory but significantly meaningful knowledge. Therefore, the innovation of this study is to incorporate ARM techniques with an exceptionality measurement to simultaneously extract common and exception rules for revealing thyroid disease associations. The contributions can be summarized as follows:

- This study proposes a novel framework which incorporates text mining procedures with ARM techniques for efficient rule extraction from raw digital health records.
- An exceptionality measurement was involved to extract common and exception rules simultaneously, conducting to enhanced efficiency of in-depth knowledge discovery.
- Two distinctive digital health datasets were analyzed through two ARM algorithms. The mutual rules were selected and interpreted. The proposed framework exhibits sound generalization, indicating its potential feasibility in other disease associations identification task.
- This study reveals the associations of thyroid disease so that the identified risk factors can be countered to enlarge the use of precision medicine while alleviating morbidity and mortality rates.
- This work is reproducible as the source code and datasets can be approached via the GitHub link <https://github.com/Amyyy-z/Association-rule-mining>.

2. Related Works

Thyroid disease has gained global attention as a result of its significant rise in incidence rates. Researchers tend to focus on improving the diagnostic efficiency of thyroid disease through machine learning techniques, whereas understanding the epidemiology has been crucially ignored. This section thereupon interprets the related literature studies around thyroid disease risk factors and the application of ARM in the medical domain.

2.1. Thyroid Disease Associations Identification

Existing studies majorly focused on determining the risk factors of thyroid disease through qualitative approaches. For instance, Peterson et al. [14] has conducted a systematic literature review on 37 research studies. Their work assessed body mass index (BMI), diet habits, and female reproductive factors, and the results indicated that BMI was

associated with high risks of thyroid cancer. Similarly, Shih et al. [15] summarized existing studies and put their effort into investigating the associations between diabetes and thyroid cancer. Their results suggested that the association between diabetes and thyroid cancer was relatively weak, and further experiments were required for confirmation. Additionally, more studies proposed risk factors correlated with thyroid disease through such a qualitative approach, such as vitamin D deficiency [16], obesity [17,18], radiation [19], hormonal factors [20], and gene mutations [4,21].

Nevertheless, many of the aforementioned factors are still controversial as they were not empirically and experimentally evaluated. Among all those literature studies, only a few studies adopted case-control and retrospective approaches to evaluate risk factors of thyroid disease [22–25]. However, these studies generally investigate one factor at a time. The evaluations of either one of those factors take long-established retrospective investigations, let alone the inner connections among them were ignored. More importantly, those studies were usually conducted with different groups of patients under diversified demographic features and scales. Therefore, no consensus can be established on the investigated factors. To address such a challenge and to verify the controversial risk factors, this study bridges the literature gaps and adopts ARM techniques inclusive of exception rule measures to investigate the associations of thyroid disease.

2.2. Association Rule Mining in Medical Domain

ARM has been deployed in many research fields, including time series analysis [26], retailing pattern analysis [27], and educational technologies performance [28]. In addition, the medical field is another intense domain that has adopted ARM frequently for knowledge discovery [29–33].

For instance, Lee and Cartmell [34] has investigated 897 cancer survivors. Through the use of ARM algorithms, they found a significantly higher risk of current smoking status with cardiovascular disease development. Likewise, Cha and Kim [11] applied the Apriori algorithm on 7709 patients with mood disorders to identify the potential comorbidities, and their results suggested that diabetes and hypertension were strongly related to mood disorders. Umasankar and Thiagarasu [35] adopted ARM techniques to identify criteria that might lead to heart attacks. They found strong relationships among “chest pain, thallium scan, exercise-induced angina, and major coloured vessels”. Yavari et al. [36] proposed to use fuzzy ARM on detecting heart disease, to name a few. Many more studies demonstrated that ARM is effective in identifying underlying correlations among comorbidity [9,37–39].

Moreover, existing studies mainly focused on identifying common rules for a particular disease, whereas researchers have considerably neglected exception rule extraction. Taniar et al. [12] stated that exception rules also produce helpful knowledge and should not be overlooked. In fact, they may be even more valuable than common rules, particularly within the medical domain, due to their ability to provide actionable insights that hold significant importance. More specifically, people can be of primary control to avoid certain exceptional factors to mitigate the possibility of being diagnosed with a particular disease or even enlarge the potential use of precision medicine in clinical practice, more importantly, to alleviate disease mortality rates, extend survival, and improve quality of life [40]. Hence, this study integrates text mining procedures, ARM techniques, and exceptionality measurement to simultaneously identify common and reliable exception rules from digital health records, enlarging society’s benefits.

3. Materials and Methods

To better illustrate the research design, this section presents the proposed framework, the adopted ARM algorithms, the utilized exceptionality measurements, and the algorithm for exception rules generation.

3.1. Proposed TM-ARM Framework

Figure 1 presents the proposed TM-ARM framework, which incorporates text mining, association rule mining, and exceptionality identification. The raw records were obtained

from electronic health systems containing the de-identified patient's admission reports and discharge summaries. Those documents include demographic features, medical history, comorbidity, lifestyle behaviors, diagnostic reports, symptoms, disease stage, treatment protocols, and principal diagnosis. Text mining procedures were involved here to extract key risk factors from raw health records. After breaking the clinical notes and reports into tokens, only the medical terminologies were extracted and normalized through stemming and lemmatization, and a set of stop-words were defined and removed.

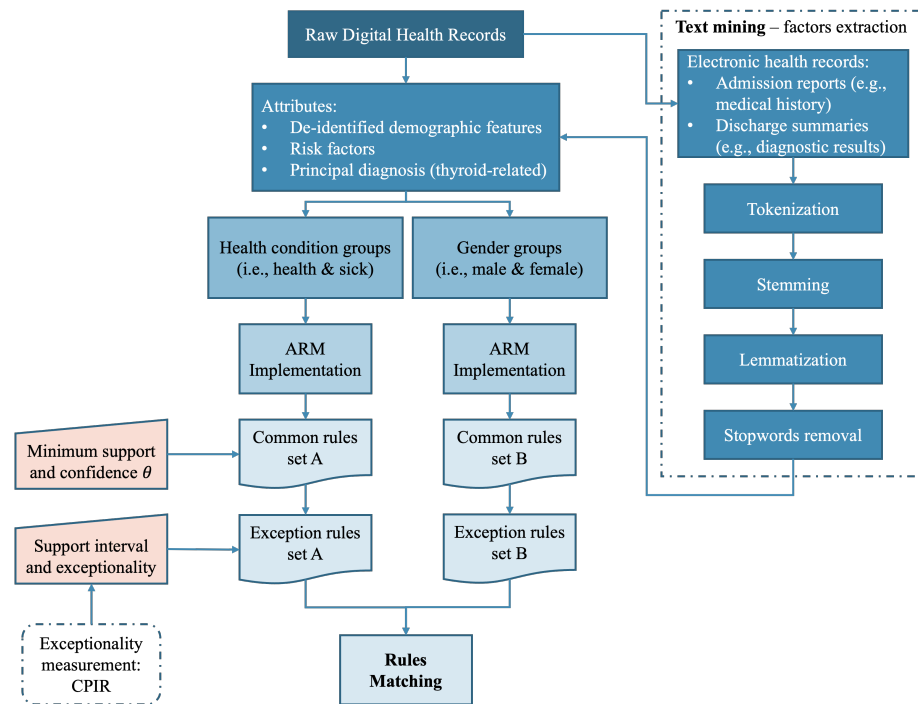


Figure 1. Text mining—Association rule mining with exceptional measurement integration (TM-ARM) framework.

In order to achieve comprehensive investigations, two health datasets were adopted in this study so that the extracted rules from two independent sources can be analyzed to ensure the fairness and reliability of the findings. More specifically, the open-access thyroid-related dataset was retrieved from the UCI data repository [41], which has already been pre-processed. This dataset was used to reveal risk factors correlated with functional thyroid disease. The self-acquired dataset was obtained from a first-class hospital in China which requires the text mining procedure to extract critical attributes from the electronic health system. This dataset was de-identified and re-labelled to identify risk factors associated with neoplastic thyroid disease.

Both datasets were split into different groups for precise rules extraction, including healthy condition groups (i.e., healthy and sick) and gender groups (i.e., male and female). Each group was evaluated with two classic ARM algorithms. The exception rules will also be generated simultaneously with the involvement of exceptionality measurement. During the analysis phase, the conflicting rules from two ARM algorithms were removed following the forward and backward chaining reasoning [42], where the distinct rules were excluded and only retained the mutual rules. The final mutual rules generated from the two algorithms were regarded as the risk factors associated with thyroid disease.

3.2. Association Rule Mining Algorithms

ARM was introduced by Agrawal et al. [43] to discover the occurrence of items in market transactions. The basic concept is to verify $X \Rightarrow Y$ indicating that if an item X exist, then item Y should co-exist. In order to identify the correlations, let $D = \{X_1, X_2, \dots, X_n\}$, where D is the original database, X_i is the i th instance in D , and n is the total number of

instances in D . For each X_i , there might be k items, in other words, k number of attributes or risk factors. To generate frequent itemsets, the support value of each item in X_i will be included and calculated through Equation (1), which is to identify the frequency of an itemset.

$$Support = \frac{freq(X \Rightarrow Y)}{n} \quad (1)$$

When evaluating the frequent itemsets, the conditional probability $P(Y|X)$ will also need to be paid attention to through Equation (2) to identify the confidence that an instance contains X and also contains Y .

$$Confidence = \frac{freq(X \Rightarrow Y)}{freq(X)} \quad (2)$$

In the contemporary, ARM was utilized in diverse research domains with increasing frequency. In this study, the two most classic ARM algorithms will be adopted for thyroid disease risk factor generation, which are the Apriori algorithm and the FP-Tree algorithm.

3.2.1. Apriori

The Apriori algorithm was initially presented by Agrawal and Srikant [44], and the goal is to extract associations, frequent patterns, or even casual structures from unstructured datasets. The algorithm is relatively straightforward for implementation, requiring the identification of candidate itemsets for frequent rule sets generation. The detailed procedures are as follows:

- Pre-define thresholds for support and confidence values.
- Identify support values for all individual items k in X_i , then prune the ones which do not meet support threshold.
- Loop through iteration in D , for each candidate k item in X_i , pare it up until enumerating all items in X_i .
- Calculate support values for all the candidate itemsets and prune the ones below the threshold.
- Repeat the above two steps, each time including itemsets $k + 1$ for each X_i , until finishing listing all itemsets in D .
- Final rules are the frequent itemsets with support and confidence above thresholds.

3.2.2. FP-Tree

The frequent pattern-growth tree (FP-Tree) was proposed by Han et al. [45]. It is another classic ARM algorithm adopted relatively often in the medical domain. Unlike Apriori, the FP-Tree algorithm does not require the generation of candidate itemsets, making the rules extraction process more efficient when dealing with small-to-medium scaled datasets. The detailed procedures are as follows:

- Identify support values for all individual k items in X_i .
- Write all the items k in descending order based on the support values.
- Draw the FP-tree starting from the “null” node and record the k items following the descending list.
- Update the generated tree through each iteration; meanwhile, record and update the item frequency in the tree structure.
- Generate a conditional FP-tree if the support for the node is larger than the threshold.
- Generate frequent patterns based on the conditional FP-tree as final rules.

3.3. Exceptionality Measurement

Exception rules have low support values but high confidence values. The most well-known exception rule is “*Champagne* \Rightarrow *Caviar*”, which generally does not have a high frequency in the database since they are pricey, but they are always brought together [12]. Exception rules can be influential and valuable. We incorporated the exceptionality mea-

surement to identify exceptional underlying knowledge from digital health records for revealing thyroid disease associations.

Based on Piatetsky–Shapiro’s arguments [46] and the probability theory, the measurements of common and exception rules should be different. Therefore, we incorporated the conditional-probability increment ratio (CPIR) function proposed by Wu et al. [47] as an additional measurement for rules selection to evaluate the dependency of the antecedent X and the consequent Y . In particular, the CPIR function for common rules evaluation is through Equation (3), and for exception rules evaluation is through Equation (4).

$$CPIR(X_i^k \Rightarrow Y_j) = \frac{sup(X_i^k \cup Y_j) - sup(X_i^k) \times sup(Y_j)}{sup(X_i^k) \times (1 - sup(Y_j))} \quad (3)$$

$$CPIR(X_i^k \Rightarrow \neg Y_j) = \frac{sup(X_i^k \cup \neg Y_j) - sup(X_i^k) \times sup(\neg Y_j)}{sup(X_i^k) \times sup(Y_j)} \quad (4)$$

This study aims to extract common and exception rules simultaneously for increased efficiency. Thus, reliable common and exception rules can be identified with the thresholds set for support, confidence, and CPIR. The detailed procedure of rules generation is interpreted in Algorithm 1.

Algorithm 1: Pseudo-code for exception rules generation

Input:

$D = \{X^1, X^2, \dots, X^n\}$; X is the patient instances and n is total number of cases
 $\theta_{Sup} = (n, m)$, θ_{Con} ; Set threshold interval for support and minimum confidence

Initialization:

ARM algorithm: Apriori or FP-Tree

Set minimum support $\theta_{Sup} = n$

$Rules \leftarrow (Sup, Con)$; Rule sets generated through ARM with minimum support

Set $i = 0$

while $i \leq len(Rules)$ **do**

if $Con_i \geq \theta_{Con}$ **then**

 | $Rules \leftarrow (Sup_i, Con_i)$; Store rules greater than minimum confidence

end

 Set $\theta_{Sup} = m$

for i in $Rules$ **do**

if $Sup_i \leq \theta_{Sup}$ **then**

 | $Rules \leftarrow (Sup_i, Con_i)$; Store rules in between support interval

end

 Set θ_{CPIR}

 Calculate CPIR for each i in $Rules$ using Equations (3) and (4)

for i in $Rules$ **do**

if $CPIR_i \geq \theta_{CPIR}$ **then**

 | $Rules \leftarrow (Sup_i, Con_i)$; Rules greater than threshold CPIR will be stored

end

end

end

$i = i + 1$

end

Store and plot all the final rules

Output: Final exception rules $\Rightarrow Exp_{final}$

4. Experimental Setup

This study follows a rigorous procedure for experimental settings. This section demonstrates the adopted datasets and the parameter setting.

4.1. Dataset Descriptions

This research includes two datasets to evaluate the proposed framework for comprehensively understanding thyroid disease associations.

4.1.1. Open-Access Dataset

The open-access dataset was retrieved from the UCI machine learning repository [41]. This dataset contains 22 attributes and 2800 instances with thyroid disease-related diagnoses. After the data wrangling process, a total number of 2689 instances with 14 attributes were utilized in this research, and the selected attributes can be found in Table 1.

Table 1. UCI dataset selected attributes.

Attributes	Descriptions
Age	Age group intervals
Sex	M = Male or F = female
Sick	Current sick status; False or True
Goiter	Have or had goiter; False or True
Tumor	Have or had tumor; False or True
Pregnant	Current pregnant status; False or True
Psych	Have or had psych disease; False or True
I-131	Have or had I-131 treatment; False or True
Thyroid surgery	Have or had thyroid surgery; False or True
Thyroxine	Take or have taken thyroxine; False or True
Query hypothyroid	Have or had hypothyroidism; False or True
Query hyperthyroid	Have or had hyperthyroidism; False or True
Anti-thyroid medication	Take or have taken anti-thyroid medication; False or True
Class	Current thyroid disease status; Negative or Positive

4.1.2. Self-Acquired Dataset

In order to conduct a comprehensive investigation of thyroid disease associations, this study also involves a self-acquired dataset for understanding factors related to thyroid cancer. We obtained 578 in-patient digital health records from a first-class Chinese hospital with ethics approval from Monash University Human Research Ethics Committee.

Those records were obtained from August 2018 to August 2021 in-patient data, including raw admission reports, diagnostic reports, and discharge summaries, in .pdf format. This study then incorporates text-mining procedures to extract critical factors from those digital health records. The extracted attributes were adopted for ARM implementation to identify whether they correlate with thyroid cancer.

The raw admission reports and discharge summaries were used for attributes extraction. The admission reports contain the patient's demographic information, medical history, lifestyle behaviors, and current symptoms. The discharge summaries include the patient's treatment protocols, comorbidities, and principal diagnosis. The extracted attributes were normalized through stemming and lemmatization procedures. In order to reach consistency, the extracted information include not only the same attributes from the UCI repository but also the history of diseases, comorbidity, and principal diagnosis. Therefore, 20 attributes were adopted from the hospital dataset, and details are available in Table 2.

Table 2. Self-acquired dataset selected attributes.

Attributes	Descriptions
Obesity	Have or had obesity; False or True
Diabetes	Have or had diabetes; False or True
Depression	Have or had depression; False or True
Radiation	Had radiation exposure; False or True
Hypertension	Have or had hypertension; False or True
Vitamin D deficiency	Have or had VD deficiency; False or True

4.2. Parameters Setting

This study follows a rigorous procedure during the experimental setup. For the open-access dataset, the final list of attributes and instances were selected based on the following mechanism:

- All the instances with missing ages were removed.
- All the instances with missing gender were removed.
- All the categorical variables were transformed into numerical values.
- Numerical variables with missing values were assigned random numbers between normal ranges of the blood examinations: TSH: 0.27–4.2, T3: 1.3–3.1, TT4: 62–164, T4U: 0.7–1.8, FTI: 53–142.

The support threshold for ARM implementation was set to 0.7, and the confidence threshold was 0.95 for common rules extraction. As for the exception rules generation, the support interval was set to (0.2, 0.4] (i.e., >0.2 and ≤ 0.4) with the same confidence threshold as common rules.

The self-acquired dataset selects attributes through the mechanism as follows:

- All the instances with the principal diagnosis missing or unclear were removed.
- Risk factors present denoted with 1, absent as 0.
- Numerical variables with missing values were assigned random numbers between normal ranges of the blood examinations: FT3: 3.6–7.5, FT4: 12–22, TGII: 3.5–77, TGAAb: 11–115, TPOAb: 0–34.

For the self-acquired dataset, the minimum support value was set lower than the open-access data settings because the scale of the private dataset was relatively small. Therefore, the minimum support value and confidence threshold for common rules were 0.6 and 0.9, respectively. As for the exception rules, the support interval was also set to (0.2, 0.4], and the confidence threshold was still 0.9. In addition, the minimum CPIR score for the UCI dataset was set to 0.50, for the hospital dataset was set to 0.2, and the final rules were sorted by confidence and CPIR values.

5. Results

The results for the two datasets are presented with the implementation of the Apriori and the FP-Tree algorithms. Since the two algorithms generated similar rules, the mutual rules from both algorithms were selected and interpreted. In addition, the conflicting rules were compared and investigated following forward and backward reasoning. The retained rules are reliable ones and are presented in this section.

5.1. Functional Thyroid Disease Associations

Table 3 presents the common and exception rules for health condition groups generated through the open-access dataset. For the common rules generated through Apriori and FP-Tree, it is evident that thyroxine status, hyperthyroidism, and tumor histories are three attributes correlated to current thyroid disease. If the patient does not have the aforementioned disease history, they are very likely to belong to the healthy group. On the contrary, the positive class is highly related to the female gender groups. When considering the exception rules generated in the health condition groups, age group plays a critical

role in differentiating healthy and sick instances. People aged from 56 to 70 will be more likely to be free from thyroid disease if they do not have a history of hyperthyroidism or goiter. However, females aged from 19 to 35 share more possibilities of being diagnosed with thyroid disease with a CPIR score over 0.8.

Table 4 demonstrates the extracted rules under different gender groups. It exhibits that male thyroid disease status tend to be associated with thyroxine status and thyroid surgery history, whereas female hyperthyroidism histories might be influential to the disease status. As for the exception rules, both male and female groups are less likely to be detected with thyroid disease if they age between 56 to 70 with no history of anti-thyroid medication intake.

Table 3. Functional thyroid disease associations (health groups)—common and exception rules.

Common Rules					
Groups	Association Rules		Class	Confidence	CPIR
Healthy	Thyroxine = F, Hyperthyroid = F, Tumor = F	⇒	Negative	1.00	0.81
	Thyroxine = F, Hyperthyroid = F, Tumor = F, I131 = F	⇒	Negative	1.00	0.81
	Thyroxine = F, Hyperthyroid = F, Tumor = F, Psych = F	⇒	Negative	0.99	0.80
	Hyperthyroid = F, Tumor = F, Sick = F	⇒	Negative	0.99	0.79
	Hyperthyroid = F, Tumor = F, Sick = F, Goiter = F	⇒	Negative	0.99	0.73
Sick	Sex = F, Hypothyroid = F	⇒	Positive	1.00	1.00
	Sex = F, Hyperthyroid = F, Thyroxine = F	⇒	Positive	1.00	1.00
	Sex = F, Hypothyroid = F, Thyroid_surgery = F	⇒	Positive	0.99	0.62
	Sex = F, Hypothyroid = F, Goiter = F	⇒	Positive	0.99	0.61
	Sex = F, Hypothyroid = F, Goiter = F, Sick = F	⇒	Positive	0.99	0.60
Exception Rules					
Healthy	Age_group = 56–70, Hyperthyroid = F, Goiter = F	⇒	Negative	1.00	1.00
	Age_group = 56–70, Hyperthyroid = F, Goiter = F, Sick = F	⇒	Negative	1.00	1.00
	Age_group = 56–70, Hyperthyroid = F, I131 = F, Pregnant = F	⇒	Negative	1.00	1.00
	Sex = M, Hyperthyroid, Sick = F, Tumor = F, Thyroid_surgery = F	⇒	Negative	1.00	1.00
	Sex = M, Hyperthyroid, Goiter = F, Psych = F, I131 = F, Thyroxine = F	⇒	Negative	1.00	1.00
Sick	Age_group = 36–55, Goiter = F	⇒	Positive	1.00	1.00
	Age_group = 19–35, Sex = F, Pregnant = F, Goiter = F, Sick = F, I131 = F	⇒	Positive	1.00	1.00
	Age_group = 19–35, Sex = F, Pregnant = F, Goiter = F, Anti-thyroid_meds = F	⇒	Positive	1.00	1.00
	Age_group = 19–35, Pregnant = F, Goiter = F, Sick = F, Psych = F	⇒	Positive	0.97	0.81
	Age_group = 19–35, Pregnant = F, Goiter = F, I131 = F, Psych = F	⇒	Positive	0.98	0.81

Table 4. Functional thyroid disease associations (gender groups)—common and exception rules.

Common Rules					
Groups	Association Rules		Class	Confidence	CPIR
Male	Sex = M, Thyroxine = F, Thyroid_surgery = F	⇒	Negative	1.00	0.64
	Sex = M, Thyroxine = F, Thyroid_surgery = F, Goiter = F	⇒	Negative	1.00	0.63
	Sex = M, Thyroxine = F, I131 = F	⇒	Negative	1.00	0.63
	Sex = M, Thyroxine = F, Thyroid_surgery = F, Tumor = F	⇒	Negative	1.00	0.62
	Sex = M, Thyroxine = F, Thyroid_surgery = F, Hyperthyroid = F	⇒	Negative	0.99	0.62
Female	Sex = F, Thyroxine = F, Pregnant = F, Hyperthyroid, I131 = F	⇒	Negative	1.00	0.93
	Sex = F, Thyroxine = F, Pregnant = F, Hyperthyroid = F, Tumor = F	⇒	Negative	1.00	0.93
	Sex = F, Pregnant = F, Hyperthyroid, Tumor = F	⇒	Negative	1.00	0.91
	Sex = F, Pregnant = F, Hyperthyroid, Tumor = F, Psych = F	⇒	Negative	0.99	0.90
	Sex = F, Pregnant = F, Thyroxine = F, Tumor = F	⇒	Negative	0.99	0.81

Table 4. Cont.

Exception Rules					
Male	Sex = M, Age_group = 56–70, Thyroid_surgery = F	⇒	Negative	1.00	0.83
	Sex = M, Age_group = 56–70, Thyroid_surgery = F, Hypothyroid = F	⇒	Negative	1.00	0.82
	Sex = M, Age_group = 56–70, Thyroid_surgery = F, Psych = F, Anti-thyroid_meds = F	⇒	Negative	1.00	0.82
	Sex = M, Age_group = 56–70, Hypothyroid = F, Goiter = F, Psych = F	⇒	Negative	1.00	0.81
	Sex = M, Age_group = 56–70, Goiter = F, Anti-thyroid_meds = F, Sick = F	⇒	Negative	0.99	0.80
Female	Sex = F, Age_group = 56–70, Anti-thyroid_meds = F, Goiter = F, Tumor = F	⇒	Negative	1.00	1.00
	Sex = F, Age_group = 56–70, Anti-thyroid_meds = F, Goiter = F, Thyroid_surgery = F	⇒	Negative	1.00	0.94
	Sex = F, Age_group = 56–70, Anti-thyroid_meds = F, Tumor = F	⇒	Negative	1.00	0.94
	Sex = F, Age_group = 56–70, Anti-thyroid_meds = F, Hyperthyroid = F	⇒	Negative	1.00	0.92
	Sex = F, Age_group = 56–70, Anti-thyroid_meds = F	⇒	Negative	0.99	0.87

5.2. Neoplastic Thyroid Disease Associations

To understand risk factors correlated with neoplastic thyroid disease, the proposed TM-ARM framework was also evaluated with the self-acquired dataset. Table 5 presents the health condition group results for the self-acquired dataset. It shows that the most frequent factors appearing in the healthy group are anti-thyroid medication intake status and history of hyperthyroidism. Nevertheless, in the sick group, females and a history of thyroid surgery and tumor are the two most critical attributes associated with thyroid disease. Regarding the exception rules, the history of hypothyroidism and hyperthyroidism play significant roles in thyroid disease status. In addition, when a person is free from diabetes and hypertension, he or she is very likely to be free from thyroid disease. The sick group indicates that thyroid disease is highly related to age and sex, and thyroid surgery history also plays a critical role.

Table 5. Neoplastic thyroid disease associations (health groups)—common and exception rules.

Common Rules					
Groups	Association Rules		Class	Confidence	CPIR
Healthy	Anti-thyroid_meds = F, Hyperthyroid = F, I131 = F	⇒	Negative	1.00	1.00
	Anti-thyroid_meds = F, Hyperthyroid = F, Sick = F, Goiter = F	⇒	Negative	1.00	0.96
	Anti-thyroid_meds = F, Hyperthyroid = F, I131 = F, Psych = F	⇒	Negative	0.99	0.90
	Hyperthyroid = F, I131 = F, Depression = F, Pregnant = F, Obesity = F	⇒	Negative	0.99	0.79
	Depression = F, Vitamin_D_Deficiency = F, Psych = F	⇒	Negative	0.99	0.71
Sick	Sex = F, Thyroid_surgery = T, Vitamin_D_Deficiency = F	⇒	Positive	1.00	1.00
	Sex = F, Thyroid_surgery = T, Depression = F, Anti-thyroid_meds = F	⇒	Positive	1.00	1.00
	Sex = F, Thyroid_surgery = T, Radiation = F, Obesity = F	⇒	Positive	1.00	1.00
	Sex = F, Thyroid_surgery = T, Vitamin_D_Def = F, Obesity = F, Tumor = T	⇒	Positive	1.00	1.00
	Sex = F, Thyroid_surgery = T, Radiation = F, Obesity = F, Tumor = F	⇒	Positive	1.00	1.00
Exception Rules					
Healthy	Hypothyroid = F, Thyroxine = F, Tumor = F	⇒	Negative	1.00	0.20
	Hypothyroid = F, Thyroxine = F, Tumor = F, Thyroid_surgery = F	⇒	Negative	1.00	0.20
	Hyperthyroid = F, Thyroxine = F, Diabetes = F, Hypertension = F	⇒	Negative	1.00	0.20
	Hyperthyroid = F, Diabetes = F, Hypertension = F, Depress = F	⇒	Negative	1.00	0.20
	Hyperthyroid = F, Diabetes = F, Hypertension = F, Obesity = F	⇒	Negative	1.00	0.20
Sick	Age_group = 56–70, Sex = F, Thyroid_surgery = T, Tumor = T	⇒	Positive	1.00	1.00
	Age_group = 56–70, Sex = F, Thyroid_surgery = T, Hypothyroid = T	⇒	Positive	1.00	1.00
	Age_group = 19–35, Sex = F, Thyroxine = T, Thyroid_surgery = T, Tumor = T	⇒	Positive	1.00	1.00
	Age_group = 19–35, Sex = F, Thyroxine = T, Thyroid_surgery = T, Tumor = T	⇒	Positive	1.00	1.00
	Age_group = 19–35, Thyroid_surgery = T, Thyroxine = T, Tumor = T, I131 = T	⇒	Positive	1.00	1.00

Additionally, in the gender group results, Table 6 shows that if a male patient does not have a history of thyroid surgery, radiation, or depression, he will be less likely to obtain thyroid-related diseases. On the other hand, for the female groups, Iodine-131

treatment history, pregnancy status, and obesity status are three primary factors correlated with thyroid disease. The patterns shown for the exception rules are intriguing, showing that diabetes and anti-thyroid medication status are associated with thyroid disease for the male group. For the female group, a history of hyperthyroidism, hypertension, and depression is critical to thyroid disease existence. Additionally, vitamin D deficiency is relatively essential to the establishment of thyroid disease. If a female is free from those factors, she is very likely to be safe from thyroid disease.

Table 6. Neoplastic thyroid disease associations (gender groups)—common and exception rules.

Common Rules					
Groups	Association Rules		Class	Confidence	CPIR
Male	Sex = M, Thyroid_surgery = F	⇒	Negative	0.99	0.72
	Sex = M, Thyroid_surgery = F, Radiation = F	⇒	Negative	0.99	0.72
	Sex = M, Thyroid_surgery = F, Depression = F, Psych = F	⇒	Negative	0.99	0.71
	Sex = M, Radiation = F, Depression = F, Goiter = F	⇒	Negative	0.99	0.70
	Sex = M, Radiation = F, Depression = F, Psych = F, Obesity = F	⇒	Negative	0.98	0.69
Female	Sex = F, I131 = F, Radiation = F	⇒	Negative	0.96	0.51
	Sex = F, I131 = F, Depression = F	⇒	Negative	0.96	0.50
	Sex = F, I131 = F, Pregnant = F, Psych = F	⇒	Negative	0.96	0.50
	Sex = F, I131 = F, Pregnant = F, Radiation = F, Sick = F, Obesity = F	⇒	Negative	0.96	0.50
	Sex = F, I131 = F, Pregnant = F, Psych = F, Depression = F, Obesity = F	⇒	Negative	0.96	0.50
Exception Rules					
Male	Sex = M, Diabetes = F, Sick = F, Radiation = F	⇒	Negative	1.00	0.35
	Sex = M, Anti-thyroid_meds = F, Diabetes = F, Depression = F, Thyroid_surgery = F	⇒	Negative	1.00	0.35
	Sex = M, Anti-thyroid_meds = F, Diabetes = F, Sick = F, Depression = F	⇒	Negative	1.00	0.35
	Sex = M, Anti-thyroid_meds = F, Diabetes = F, Sick = F, Depression = F, I131 = F	⇒	Negative	1.00	0.33
	Sex = M, Anti-thyroid_meds = F, Diabetes = F, Sick = F, Thyroid_surgery = F	⇒	Negative	1.00	0.31
Female	Sex = F, Hyperthyroid = F, Hypertension = F, Obesity = F, Depression = F	⇒	Negative	1.00	0.20
	Sex = F, Hyperthyroid = F, Hypertension = F, Goiter = F, Thyroid_surgery = F	⇒	Negative	1.00	0.20
	Sex = F, Hyperthyroid = F, VD_Deficiency = F, Tumor = F, Thyroid_surgery = F	⇒	Negative	1.00	0.20
	Sex = F, Hyperthyroid = F, Obesity = F, Diabetes = F, Depression = F	⇒	Negative	1.00	0.20
	Sex = F, Hyperthyroid = F, VD_Deficiency = F, Tumor = F, Depression = F	⇒	Negative	1.00	0.20

6. Discussion

Although thyroid disease is prevalent today, the cause of it remains unclear. ARM techniques have been applied relatively often in the medical domain, whereas generating rules directly from raw digital health records and identifying exception rules are the novelties in this study. Therefore, this study proposed an integrated framework consisting of text mining procedures, ARM algorithms, and exceptionality measurement to identify the associations of thyroid disease through two independent digital health records.

Based on evaluating two datasets, we confirmed that gender and age are the two leading factors correlated to thyroid disease, and this finding aligns with the existing works [3]. The reason behind this might be due to hormonal factors, including the effects of pregnancy or pubertal development, and these are particularly sensitive to young females. From the result patterns, it is quite evident that females aged from 19 to 35 are highly exposed to thyroid disease, especially thyroid cancer. And this finding is concordant with the report proposed by Australian Institute of Health and Welfare that thyroid cancer was the commonest cancer diagnosis in women aged between 25 and 29 [48].

In addition to age and gender, the results generated from both datasets manifest that a history of thyroid-related diseases, such as hypothyroidism, hyperthyroidism, goiter, or past thyroid surgery, increases the risk of subsequent thyroid diseases. This finding is intriguing and matches with the existing study proposed by Jackson et al. [49]. This finding confirmed that a subsequent thyroid cancer risk was highly enhanced if there had been an existence of thyroid disease in the past. Therefore, a history of thyroid-related diseases can be a good indicator when diagnosing current thyroid status.

Comorbidities like diabetes, obesity, hypertension, depression, psychiatric diseases, and vitamin D deficiency were included in the self-acquired dataset for evaluation. Among all the factors, psychiatric diseases are not related to the associations of neoplastic thyroid disease. For the other factors, the results exhibit no solid positive associations found between the comorbidities and thyroid disease. This finding was in accordance with [15], but controversial with [50]. Nevertheless, we found that the absence of those underlying health problems like obesity, depression, hypertension, and diabetes will reduce the risk of establishing subsequent thyroid cancer. In addition, vitamin D deficiency might be influential to thyroid disease, and this finding aligns with [24], however, further evaluations should be involved to ascertain the associations since the sample scale is relatively limited in this study. Details can be found in Figure 2, where the dash-dotted line represents the bidirectional relationship among factors.

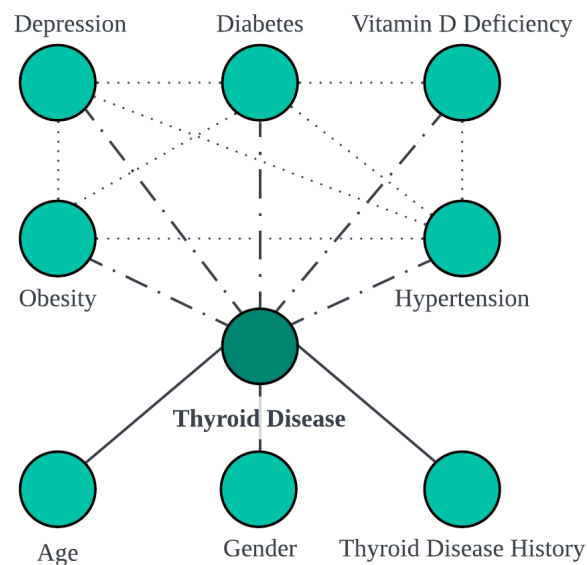


Figure 2. Final rules for thyroid disease associations.

Collectively, the proposed TM-ARM framework demonstrates solid logic and exhibits comprehensive evaluation criteria for common and exception rules generation incorporating text mining procedures. We selected thyroid disease associations for evaluation using two independent digital health records, and the extracted rules ascertained some potential risk factors associated with the disease, which also strengthened the feasibility of the framework. The implementation procedures prove that the proposed TM-ARM framework can be generalized to different diseases. However, there exist several limitations in this study. The first one is with the collected raw data from both open-access and self-acquired sources, which were initially not collected for data mining purpose, thus may not be complete. And due to the limited sample size, there might demonstrate associations which does not necessarily mean either causation. Future implementations are expected to include more samples for a more comprehensive analysis. The second limitation is the risk factors identified from the literature surveys, like gene heredity and mutations, which were not included in this study, and this will also be an alternative research direction. Moreover, as part of our future plans, we intend to enroll a larger cohort of patients to further validate the proposed method and the findings obtained in this study.

7. Conclusions

To conclude, although ARM has been famous for the last few decades, the extraction of exception rules has been considerably neglected. In addition, manually mining from heterogeneous health records requires great effort, which is impracticable. Accordingly, this study proposes a TM-ARM framework that integrates text mining procedures and

association rule mining with exceptionality, contributing to mining valuable knowledge around thyroid disease associations. Moreover, the proposed framework was analyzed through two independent digital health records, in which conflicting rules were excluded following the forward and backward reasoning and only retained the mutual rules.

The findings suggest that age and gender are the two most critical factors associated with thyroid disease. In addition, past thyroid-related disease history will increase the risk of subsequent thyroid cancer establishment. In addition, the absence of depression, vitamin D deficiency, diabetes, hypertension, or obesity are associated with reduced likelihood of thyroid disease development. Furthermore, the proposed TM-ARM framework can be expanded to encompass different diseases, facilitating in-depth knowledge discovery. This extension has the potential to contribute to the prevention of specific diseases by identifying its associations.

Author Contributions: Conceptualization, X.Z. and V.C.S.L.; methodology, X.Z. and V.C.L.; software, X.Z.; validation, X.Z., V.C.S.L. and J.C.L.; formal analysis, X.Z., V.C.S.L. and J.C.L.; investigation, X.Z.; resources, X.Z.; data curation, X.Z.; writing—original draft preparation, X.Z. and V.C.L.; writing—review and editing, X.Z., V.C.S.L. and J.C.L.; visualization, X.Z., V.C.S.L. and J.C.L.; supervision, V.C.S.L. and J.C.L.; project administration, V.C.S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki, and approved by the Human Ethics Committee of Monash University (Project ID: 24704 approved on 27 July 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: This study provides partial research data to be analyzed due to privacy issues, and the dataset can be found in the link <https://github.com/Amyyy-z/Association-rule-mining>.

Acknowledgments: The authors would like to thank the participant institution. We thank Feng Liu, Jackie Rong, Qilin Zhang, and Haoyu Kong for their discussions and constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Apostu, D.; Lucaciu, O.; Oltean-Dan, D.; Mureşan, A.D.; Moisescu-Pop, C.; Maxim, A.; Benea, H. The influence of thyroid pathology on osteoporosis and fracture risk: A review. *Diagnostics* **2020**, *10*, 149. [[CrossRef](#)] [[PubMed](#)]
2. Rao, A.; Renuka, B. A Machine Learning Approach to Predict Thyroid Disease at Early Stages of Diagnosis. In Proceedings of the 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangalore, India, 6–8 November 2020; pp. 1–4. [[CrossRef](#)]
3. Society, A.C. Cancer Facts & Figures 2023. American Cancer Society. Available online: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2023/2023-cancer-facts-and-figures.pdf> (accessed on 20 June 2023).
4. Efanov, A.A.; Brenner, A.V.; Bogdanova, T.I.; Kelly, L.M.; Liu, P.; Little, M.P.; Wald, A.I.; Hatch, M.; Zurnadzy, L.Y.; Nikiforova, M.N.; et al. Investigation of the Relationship Between Radiation Dose and Gene Mutations and Fusions in Post-Chernobyl Thyroid Cancer. *JNCI J. Natl. Cancer Inst.* **2017**, *110*, 371–378. [[CrossRef](#)] [[PubMed](#)]
5. Fiore, M.; Oliveri Conti, G.; Caltabiano, R.; Buffone, A.; Zuccarello, P.; Cormaci, L.; Cannizzaro, M.A.; Ferrante, M. Role of Emerging Environmental Risk Factors in Thyroid Cancer: A Brief Review. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1185. [[CrossRef](#)] [[PubMed](#)]
6. Egalini, F.; Parasiliti Caprino, M.; Gaggero, G.; Cappiello, V.; Giannelli, J.; Rossetto Giaccherino, R.; Pagano, L.; Giordano, R. Endocrine Disorders in Autoimmune Rheumatological Diseases: A Focus on Thyroid Autoimmune Diseases and on the Effects of Chronic Glucocorticoid Treatment. *Endocrines* **2021**, *2*, 171–184. [[CrossRef](#)]
7. Gavryutina, I.; Fordjour, L.; Chin, V.L. Genetics of Thyroid Disorders. *Endocrines* **2022**, *3*, 198–213. [[CrossRef](#)]
8. Maciejewski, A.; Lacka, K. Vitamin D-Related Genes and Thyroid Cancer: A Systematic Review. *Int. J. Mol. Sci.* **2022**, *23*, 13661. [[CrossRef](#)] [[PubMed](#)]
9. Tandan, M.; Acharya, Y.; Pokharel, S.; Timilsina, M. Discovering symptom patterns of COVID-19 patients using association rule mining. *Comput. Biol. Med.* **2021**, *131*, 104249. [[CrossRef](#)]
10. Kadi, I.; Idri, A.; Fernandez-Aleman, J. Knowledge discovery in cardiology: A systematic literature review. *Int. J. Med. Inform.* **2017**, *97*, 12–32. [[CrossRef](#)]

11. Cha, S.; Kim, S.S. Comorbidity Patterns of Mood Disorders in Adult Inpatients: Applying Association Rule Mining. *Healthcare* **2021**, *9*, 1155. [[CrossRef](#)]
12. Taniar, D.; Rahayu, W.; Lee, V.; Daly, O. Exception rules in association rule mining. Special Issue on Advanced Intelligent Computing Theory and Methodology in Applied Mathematics and Computation. *Appl. Math. Comput.* **2008**, *205*, 735–750. [[CrossRef](#)]
13. Liu, H.; Lu, H.; Feng, L.; Hussain, F. Efficient Search of Reliable Exceptions. In *Proceedings of the Methodologies for Knowledge Discovery and Data Mining*; Zhong, N., Zhou, L., Eds.; Springer: Berlin/Heidelberg, Germany, 1999; pp. 194–204.
14. Peterson, E.; De, P.; Nuttall, R. BMI, Diet and Female Reproductive Factors as Risks for Thyroid Cancer: A Systematic Review. *PLoS ONE* **2012**, *7*, e29177. [[CrossRef](#)] [[PubMed](#)]
15. Shih, S.R.; Chiu, W.Y.; Chang, T.C.; Tseng, C.H. Diabetes and thyroid cancer risk: Literature review. *Exp. Diabetes Res.* **2012**, *2012*, 1–7. [[CrossRef](#)]
16. Zhao, R.; Zhang, W.; Ma, C.; Zhao, Y.; Xiong, R.; Wang, H.; Chen, W.; Zheng, S.G. Immunomodulatory function of vitamin D and its role in autoimmune thyroid disease. *Front. Immunol.* **2021**, *12*, 574967. [[CrossRef](#)]
17. Wang, B.; Song, R.; He, W.; Yao, Q.; Li, Q.; Jia, X.; Zhang, J.A. Sex Differences in the Associations of Obesity with Hypothyroidism and Thyroid Autoimmunity Among Chinese Adults. *Front. Physiol.* **2018**, *9*, 01397. [[CrossRef](#)] [[PubMed](#)]
18. Song, R.h.; Wang, B.; Yao, Q.m.; Li, Q.; Jia, X.; Zhang, J.A. The Impact of Obesity on Thyroid Autoimmunity and Dysfunction: A Systematic Review and Meta-Analysis. *Front. Immunol.* **2019**, *10*, 02349. [[CrossRef](#)] [[PubMed](#)]
19. Suzuki, K.; Saenko, V.; Yamashita, S.; Mitsutake, N. Radiation-Induced Thyroid Cancers: Overview of Molecular Signatures. *Cancers* **2019**, *11*, 1290. [[CrossRef](#)]
20. Nagayama, Y. Thyroid autoimmunity and thyroid cancer—the pathogenic connection: A 2018 update. *Horm. Metab. Res.* **2018**, *50*, 922–931. [[CrossRef](#)]
21. Marotta, V.; Bifulco, M.; Vitale, M. Significance of RAS mutations in thyroid benign nodules and non-medullary thyroid cancer. *Cancers* **2021**, *13*, 3785. [[CrossRef](#)]
22. Zhang, D.; Tang, J.; Kong, D.; Cui, Q.; Wang, K.; Gong, Y.; Wu, G. Impact of gender and age on the prognosis of differentiated thyroid carcinoma: A retrospective analysis based on SEER. *Horm. Cancer* **2018**, *9*, 361–370. [[CrossRef](#)]
23. Zhao, S.; Jia, X.; Fan, X.; Zhao, L.; Pang, P.; Wang, Y.; Luo, Y.; Wang, F.; Yang, G.; Wang, X.; et al. Association of obesity with the clinicopathological features of thyroid cancer in a large, operative population: a retrospective case-control study. *Medicine* **2019**, *98*, e18213. [[CrossRef](#)]
24. Zhao, J.; Wang, H.; Zhang, Z.; Zhou, X.; Yao, J.; Zhang, R.; Liao, L.; Dong, J. Vitamin D deficiency as a risk factor for thyroid cancer: A meta-analysis of case-control studies. *Nutrition* **2019**, *57*, 5–11. [[CrossRef](#)] [[PubMed](#)]
25. Kim, K.; Cho, S.W.; Park, Y.J.; Lee, K.E.; Lee, D.W.; Park, S.K. Association between Iodine Intake, Thyroid Function, and Papillary Thyroid Cancer: A Case-Control Study. *Endocrinol. Metab.* **2021**, *36*, 1034. [[CrossRef](#)] [[PubMed](#)]
26. Dhaou, A.; Bertonecello, A.; Gourvénec, S.; Garnier, J.; Le Pennec, E. Causal and Interpretable Rules for Time Series Analysis. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Virtual Event, Singapore, 14–18 August 2021*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 2764–2772.
27. Jain, A.; Jain, S.; Merh, N. Application of Association Rule Mining in a Clothing Retail Store. In *Applied Advanced Analytics*; Laha, A.K., Ed.; Springer: Singapore, 2021; pp. 103–114.
28. Guliyev, A. Analysis of Relationship Between Learning Outcomes and Student’s Exam Results Using Association Rule Mining and Fuzzy Inference Rules. In *14th International Conference on Theory and Application of Fuzzy Systems and Soft Computing—ICAIFS-2020*; Laha, A.K., Ed.; Springer: Budva, Montenegro, 2021; Volume 1306, p. 354.
29. Lakshmi, K.; Vadivu, G. Extracting Association Rules from Medical Health Records Using Multi-Criteria Decision Analysis. *Procedia Comput. Sci.* **2017**, *115*, 290–295. [[CrossRef](#)]
30. Harahap, M.; Husein, A.M.; Aisyah, S.; Lubis1, F.R.; Wijaya, B.A. Mining association rule based on the diseases population for recommendation of medicine need. *J. Phys. Conf. Ser.* **2018**, *1007*, 012017. [[CrossRef](#)]
31. Korach, Z.T.; Yang, J.; Rossetti, S.C.; Cato, K.D.; Kang, M.J.; Knaplund, C.; Schnock, K.O.; Garcia, J.P.; Jia, H.; Schwartz, J.M.; et al. Mining clinical phrases from nursing notes to discover risk factors of patient deterioration. *Int. J. Med. Inform.* **2020**, *135*, 104053. [[CrossRef](#)] [[PubMed](#)]
32. Shrestha, A.; Zikos, D.; Fegaras, L. An annotated association mining approach for extracting and visualizing interesting clinical events. *Int. J. Med. Inform.* **2021**, *148*, 104366. [[CrossRef](#)]
33. Kaur, I.; Doja, M.; Ahmad, T. Data mining and machine learning in cancer survival research: An overview and future recommendations. *J. Biomed. Inform.* **2022**, *128*, 104026. [[CrossRef](#)]
34. Lee, S.J.; Cartmell, K.B. An Association Rule Mining Analysis of Lifestyle Behavioral Risk Factors in Cancer Survivors with High Cardiovascular Disease Risk. *J. Pers. Med.* **2021**, *11*, 366. [[CrossRef](#)]
35. Umasankar, P.; Thiagarasu, V. Decision Support System for Heart Disease Diagnosis Using Interval Vague Set and Fuzzy Association Rule Mining. In *Proceedings of the 2018 4th International Conference on Devices, Circuits and Systems (ICDCS)*, Coimbatore, India, 16–17 March 2018; pp. 223–227. [[CrossRef](#)]
36. Yavari, A.; Rajabzadeh, A.; Abdali-Mohammadi, F. Profile-based assessment of diseases affective factors using fuzzy association rule mining approach: A case study in heart diseases. *J. Biomed. Inform.* **2021**, *116*, 103695. [[CrossRef](#)]

37. Peng, M.; Sundararajan, V.; Williamson, T.; Minty, E.P.; Smith, T.C.; Doktorchik, C.T.; Quan, H. Exploration of association rule mining for coding consistency and completeness assessment in inpatient administrative health data. *J. Biomed. Inform.* **2018**, *79*, 41–47. [CrossRef]
38. Jamsheela, O. Analysis of association among various attributes in medical data of heart patients by using data mining methods. *Int. J. Appl. Sci. Eng.* **2021**, *18*, 2020215. [CrossRef]
39. Ma, F.; Ye, M.; Luo, J.; Xiao, C.; Sun, J. Advances in Mining Heterogeneous Healthcare Data. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Virtual, 14–18 August 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 4050–4051.
40. Kaidar-Person, O.; Gil, Z.; Billan, S. Precision medicine in head and neck cancer. *Drug Resist. Updat.* **2018**, *40*, 13–16. [CrossRef] [PubMed]
41. Dua, D.; Graff, C. UCI Machine Learning Repository. 2017. Available online: <http://archive.ics.uci.edu/ml> (accessed on 20 June 2023).
42. Kapoor, N.; Bahl, N. Comparative study of forward and backward chaining in artificial intelligence. *Int. J. Eng. Comput. Sci.* **2016**, *5*, 16239–16242. [CrossRef]
43. Agrawal, R.; Imieliński, T.; Swami, A. Mining Association Rules between Sets of Items in Large Databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, 25–28 May 1993; Association for Computing Machinery: New York, NY, USA, 1993; pp. 207–216.
44. Agrawal, R.; Srikant, R. Fast algorithms for mining association rules. In Proceedings of the 20th International Conference on Very Large Data Bases, Santiago, Chile, 12–15 September 1994; Laha, A.K., Ed.; VLDB: Santiago de Chile, Chile, 1994; Volume 1215, pp. 487–499.
45. Han, J.; Pei, J.; Yin, Y. Mining frequent patterns without candidate generation. *ACM Sigmod Rec.* **2000**, *29*, 1–12. [CrossRef]
46. Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. From data mining to knowledge discovery in databases. *AI Magazine*, 15 March 1996.
47. Wu, X.; Zhang, C.; Zhang, S. Efficient mining of both positive and negative association rules. *ACM Trans. Inf. Syst. (TOIS)* **2004**, *22*, 381–405. [CrossRef]
48. Australian Institute of Health and Welfare. Cancer in Australia 2017. Available online: <https://www.aihw.gov.au/getmedia/3da1f3c2-30f0-4475-8aed-1f19f8e16d48/20066-cancer-2017.pdf.aspx?inline=true> (accessed on 19 June 2023).
49. Jackson, D.; Handelsman, R.S.; Farrá, J.C.; Lew, J.I. Increased incidental thyroid cancer in patients with subclinical chronic lymphocytic thyroiditis. *J. Surg. Res.* **2020**, *245*, 115–118. [CrossRef]
50. Ma, J.; Huang, M.; Wang, L.; Ye, W.; Tong, Y.; Wang, H. Obesity and risk of thyroid cancer: evidence from a meta-analysis of 21 observational studies. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **2015**, *21*, 283–291. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.