

Review

Machine learning on thyroid disease: a review

Kwang-Sig Lee^{1,*}, Hyuntae Park^{2,*}¹AI Center, Korea University College of Medicine, 02841 Seoul, Republic of Korea²Department of Obstetrics & Gynecology, Korea University College of Medicine, 02841 Seoul, Republic of Korea*Correspondence: ecophy@hanmail.net (Kwang-Sig Lee); cyberpelvis@gmail.com (Hyuntae Park)

Academic Editor: Alessandro Poggi

Submitted: 7 December 2021 Revised: 8 February 2022 Accepted: 9 February 2022 Published: 16 March 2022

Abstract

This study reviews the recent progress of machine learning for the early diagnosis of thyroid disease. Based on the results of this review, different machine learning methods would be appropriate for different types of data for the early diagnosis of thyroid disease: (1) the random forest and gradient boosting in the case of numeric data; (2) the random forest in the case of genomic data; (3) the random forest and the ensemble in the case of radiomic data; and (4) the random forest in the case of ultrasound data. Their performance measures varied within 64.3–99.5 for accuracy, 66.8–90.1 for sensitivity, 61.8–85.5 for specificity, and 64.0–96.9 for the area under the receiver operating characteristic curve. According to the findings of this review, indeed, the following attributes would be important variables for the early diagnosis of thyroid disease: clinical stage, marital status, histological type, age, nerve injury symptom, economic income, surgery type [the quality of life 3 months after thyroid cancer surgery]; tumor diameter, symptoms, extrathyroidal extension [the local recurrence of differentiated thyroid carcinoma]; RNA features including ADD3-AS1 (downregulation), MIR100HG (downregulation), FAM95C (downregulation), MORC2-AS1 (downregulation), LINC00506 (downregulation), ST7-AS1 (downregulation), LOC339059 (downregulation), MIR181A2HG (upregulation), FAM181A-AS1 (downregulation), LBX2-AS1 (upregulation), BLACAT1 (upregulation), hsa-miR-9-5p (downregulation), hsa-miR-146b-3p (upregulation), hsa-miR-199b-5p (downregulation), hsa-miR-4709-3p (upregulation), hsa-miR-34a-5p (upregulation), hsa-miR-214-3p (downregulation) [papillary thyroid carcinoma]; gut microbiota RNA features such as veillonella, paraprevotella, neisseria, rheinheimera [hypothyroidism]; and ultrasound features, i.e., wreath-shaped feature, microcalcification, strain ratio [the malignancy of thyroid nodules].

Keywords: thyroid; early diagnosis; machine learning; random forest; review

1. Introduction

The thyroid gland is an endocrine gland creating thyroid hormone. It is shaped like a butterfly and positioned in the front of the neck. Thyroid hormone involves the regulation of metabolism and various problems can occur in the gland. It can create either too little or too much hormone (hypothyroidism or hyperthyroidism). The former condition causes fatigue, weight gain and intolerance to cold temperature, whereas the latter leads to anxiety, weight loss and sensitivity to heat. Also, malignant cells can develop there (thyroid cancer) [1,2]. These disorders, thyroid disease, has been a leading cause of disease burden in the world [3–6]. The number of individuals with thyroid disease is estimated to be 200 million in the world [3], whereas the incidence and mortality of thyroid cancer registered rapid growths of 169% and 87% during 1990–2017, i.e., 95,030 and 22,070 to 255,490 and 41,240, respectively [4]. Hypothyroidism is reported to cause significant disease burden and direct, morbidity and mortality cost, as well [5,6]. It has various risk factors and many of them are still unknown. Its diagnosis and prognosis are considered to be quite challenging given that its symptoms are very similar with other diseases such as depression [1–3]. It is not surprising that there exists a high degree of variation among clinical experts in terms

of its diagnosis and prognosis. In this context, more research is to be done on this important topic. Recently, on the other hand, the terms “deep learning”, “machine learning” and “artificial intelligence” have attracted great attention all over the globe. For instance, their Google trends recorded ten-fold expansions from 10 to 100 during 2013–2018. Artificial intelligence can be defined as “the capability of a machine to imitate intelligent human behavior” (the Merriam-Webster dictionary). The definition of machine learning can be a division of artificial intelligence to “extract knowledge from large amounts of data” [7].

Six common machine learning algorithms are the decision tree, the naïve Bayesian predictor, the random forest, the support vector machine, the artificial neural network, and the deep neural network (deep learning). A decision tree has three components: an intermediate node (a test on an independent variable), a branch (an outcome of the test) and a terminal node (a value of the dependent variable). A naïve Bayesian predictor makes an early diagnosis based on Bayes’ theorem, which states that the probability of the dependent variable given certain values of independent variables comes from the probabilities of the independent variables given a certain value of the dependent variable. A random forest is a collection of many decision trees with a majority vote on the dependent variable (“bootstrap ag-



gregation”). Let us take a random forest with 1000 decision trees as an example. Here, the algorithm samples 1000 training sets with replacements, trains 1000 decision trees with the 1000 training sets, makes 1000 predictions with the 1000 decision trees, and takes a majority vote on the dependent variable. A support vector machine originates a line or space called a “hyperplane” (a collection of “support vectors”). The hyperplane divides data with the greatest distance between different sub-groups [7].

An artificial neural network is a network of “neurons”, i.e., information units combined through weights. Usually, the artificial neural network has one input layer, one, two or three intermediate layers and one output layer. Neurons in a previous layer connect with “weights” in the next layer and these weights represent the strengths of connections between neurons in a previous layer and their next-layer counterparts. This process starts from the input layer, continues through intermediate layers and ends in the output layer (feedforward operation). Then, learning happens: these weights are accommodated based on how much they contributed to the loss, a difference between the actual and predicted final outputs. This process starts from the output layer, continues through intermediate layers and ends in the input layer (backpropagation operation). The two operations are replicated until a certain expectation is met regarding the accurate diagnosis of the dependent variable. In other words, the performance of the artificial neural network improves as long as its learning continues. Finally, a deep neural network is an artificial neural network with a large number of intermediate layers, e.g., 5, 10 or even 1000. The deep neural network is called “deep learning” given that learning “deepens” through numerous intermediate layers [8].

Traditional research considers a limited scope of predictors for the early diagnosis of disease, whereas adopting logistic regression with an unrealistic assumption of *ceteris paribus*, i.e., “all the other variables staying constant”. In this context, emerging literature uses artificial intelligence for the early diagnosis of disease, e.g., arrhythmia [8], birth outcome [9–14], cancer [15–19], comorbidity [20–22], menopause [23] and temporomandibular disease [24,25]. It does not require unrealistic assumptions of “all the other variables staying constant” while managing to analyze which predictors are more important for the early diagnosis of the dependent variable. The purpose of this study is to review the recent progress of machine learning for the early diagnosis of thyroid disease.

2. Materials and methods

Twenty original studies were selected for review out of 33 original studies in PubMed with the search terms “thyroid” (title) and “random forest” (abstract). The inclusion criteria of this review were: (1) the intervention(s) of the decision tree, the naïve Bayesian predictor, the random forest, the support vector machine and/or the artificial neural

network; (2) the outcome(s) of accuracy and/or the area under the receiver operating characteristic curve for the early diagnosis of thyroid disease; (3) the publication year of 2020 or later; and (4) the publication language of English. The following summary measures were adopted: machine learning methods, sample size, data type, performance measures and important attributes (predictors). Here, accuracy can be defined as the proportion of correct predictions over all observations, while the area under the receiver operating characteristic curve (AUC) can be defined as the area under the plot of the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. The exclusion criteria of this review was that thyroid disease is an independent variable (attribute) instead of the dependent variable.

3. Results

3.1 Summary of review

The summary of review is shown in Tables 1,2 (Ref. [26–45]). The tables have five summary measures, i.e., machine learning methods, sample size, data type, performance measures, important attributes, and whether the variable importance of the random forest is reported (VI-Yes 1). Based on the results of this review, different machine learning methods would be appropriate for different types of data for the early diagnosis of thyroid disease: (1) the random forest and gradient boosting in the case of numeric data; (2) the random forest in the case of genomic data; (3) the random forest and the ensemble in the case of radiomic data; and (4) the random forest in the case of ultrasound data. Their performance measures varied within 64.3–99.5 for accuracy, 66.8–90.1 for sensitivity, 61.8–85.5 for specificity, and 64.0–96.9 for the AUC (Table 1). According to the findings of this review, indeed, the following attributes would be important variables for the early diagnosis of thyroid disease: clinical stage, marital status, histological type, age, nerve injury symptom, economic income, surgery type [the quality of life 3 months after thyroid cancer surgery]; tumor diameter, symptoms, extrathyroidal extension [the local recurrence of differentiated thyroid carcinoma]; RNA features including ADD3-AS1 (downregulation), MIR100HG (downregulation), FAM95C (downregulation), MORC2-AS1 (downregulation), LINC00506 (downregulation), ST7-AS1 (downregulation), LOC339059 (downregulation), MIR181A2HG (upregulation), FAM181A-AS1 (downregulation), LBX2-AS1 (upregulation), BLACAT1 (upregulation), hsa-miR-9-5p (downregulation), hsa-miR-146b-3p (upregulation), hsa-miR-199b-5p (downregulation), hsa-miR-4709-3p (upregulation), hsa-miR-34a-5p (upregulation), hsa-miR-214-3p (downregulation) [papillary thyroid carcinoma]; gut microbiota RNA features such as veillonella, paraprevotella, neisseria, rheinheimera [hypothyroidism]; and ultrasound features, i.e., wreath-shaped feature, micro-calcification, strain ratio [the malignancy of thyroid nodules] (Table 2). However, machine learning is a

Table 1. Summary of review: methods, sample size, data type and performance measures.

ID	Methods	Sample size	Data type	Performance
[26]	Oversampling then DT RF AB	47	Spectra	Accuracy DT 75.4 RF 81.5 AB 84.6
[27]	RF	286	Numeric	Accuracy Validation 89.7
[28]	LR RF	187	Numeric	AUC RF 77.0
[29]	LR RF	355	Numeric	RF Accuracy 71.9 AUC 85.9 Sensitivity 75.5 Specificity 82.4
[30]	LR DT RF		Numeric	Accuracy 84.7–89.7
[31]	LR RF	604	Ultrasound	AUC RF 64.0
[32]	RF	1451	Numeric	AUC 71.0–81.0
[33]	RF	428	Ultrasound	Accuracy 95.0
[34]	DT RF SVM	506	Genomic	Accuracy DT 92.2/87.0 RF 99.2/99.5 SVM 99.0/98.3 (Group 1/2 Attributes)
[35]	Ensemble	109	Radiomic	AUC 96.9
[36]	LR DT NB RF ANN		EKG	Accuracy LR 95.1 DT 91.9 NB 91.9 RF 93.5 ANN 93.5
[37]	NB RF SVM ANN	218	Numeric	Accuracy NB 81.8 RF 90.9 SVM 84.1 ANN 88.6
[38]	RF	92	Genomic	Accuracy 99.4
[39]	RF	60	Radiomic	Accuracy 90.6
[40]	LR RF SVM GB ANN	1074/6928	Numeric	Accuracy LR 80.0/76.0 RF 79.0/80.0 SVM 80.0/75.0 GB 82.0/79.0 ANN 81.0/77.0 (Thyroid Peroxidase Activity/Thyroid Hormone Receptor Modulation)
[41]	RF	60	Radiomic	Accuracy 78.6 AUC 84.9
[42]	RF	1558	Ultrasound	Accuracy 96.1
[43]	RF	92	Genomic	
[44]	LR RF SVM GB ANN	177	Ultrasound	Accuracy/AUC/Sensitivity/Specificity LR 84.2/92.8/90.1/79.2 RF 86.0/93.4/86.6/85.5 SVM 84.8/92.3/89.0/81.3 GB 83.7/92.6/85.3/82.3 ANN 84.8/90.8/87.6/82.4
[45]	LR RF SVM	96	Radiomic	Accuracy 64.3 AUC 65.1 Sensitivity 66.8 Specificity 61.8

AB, adaptive boosting; ANN, artificial neural network; AUC, area under the receiver operating characteristic curve; CNN, convolutional neural network; DT, decision tree; EKG, electroglottograph; GB, Gradient Boosting; LR, Logistic Regression; NB, Naïve Bayes; RF, Random Forest; SVM, support vector machine.

data-driven method and more study is to be done for greater external validity.

3.2 Summary of selected studies

The summary of selected studies is presented in this section. The aim of a recent study [27] was to adopt machine learning and numeric data for predicting the quality of life three months after thyroid surgery. Data came from 286 participants and the attributes were European Organization for Research and Treatment of Cancer Quality of Life Ques-

tionnaire Version 3 responses. The accuracy of the random forest for the validation set was 89.7. Based on random forest variable importance, clinical stage, marital status, histological type, age, nerve injury symptom, economic income and surgery type were the most important variables for predicting the quality of life three months after thyroid surgery. Likewise, the purpose of recent research [30] was to employ machine learning and numeric data for predicting the local recurrence of differentiated thyroid carcinoma. The accu-

Table 2. Summary of review: class, important attributes and whether variable importance (VI) is reported.

ID	Class [attributes]	Important attributes	VI-Yes
[26]	16 PTC vs. 31 Papillary Micro Carcinoma [Raman Intensity]		
[27]	Quality of Life 3 Months after Thyroid Cancer Surgery [European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Version 3]	Clinical Stage, Marital Status, Histological Type, Age, Nerve Injury Symptom, Economic Income, Surgery Type	1
[28]	Thyroid Complications after Receiving Programmed Cell Death 1/Programmed Cell Death Ligand 1 Inhibitors	Opioids	
[29]	Malignancy of Indeterminate Thyroid Nodules after Fine Needle Aspiration Biopsy [Diagnostic Pathology Features]		
[30]	Local Recurrence of Differentiated Thyroid Carcinoma	Tumor Diameter, Symptoms, Extrathyroidal Extension	1
[31]	Malignancy of Indeterminate Thyroid Nodules after Fine Needle Aspiration Cytology [Diagnostic Pathology Features]		
[32]	Local Recurrence of PTC [Serum Thyroglobulin/Antithyroglobulin Features]		
[33]	Malignancy of Thyroid Nodules [Image Features Collected Based on Scale-Invariant Feature Transformation and CNN]	Group 1 Including ADD3-AS1 (Downregulation), MIR100HG (Downregulation), FAM95C (Downregulation), MORC2-AS1 (Downregulation), LINC00506 (Downregulation), ST7-AS1 (Downregulation), LOC339059 (Downregulation), MIR181A2HG (Upregulation), FAM181A-AS1 (Downregulation), LBX2-AS1 (Upregulation), BLACAT1 (Upregulation); Group 2 Including hsa-miR-9-5p (Downregulation), hsa-miR-146b-3p (Upregulation), hsa-miR-199b-5p (Downregulation), hsa-miR-4709-3p (Upregulation), hsa-miR-34a-5p (Upregulation), hsa-miR-214-3p (Downregulation)	
[34]	PTC [703 RNA Features from Cancer Genome Atlas Data]		1

Table 2. Continued.

ID	Class [attributes]	Important attributes	VI-Yes
[35]	Papillary Thyroid Cancer [Magnetic Resonance Imaging Features]		
[36]	Hyperthyroidism/Hypothyroidism [EGG Features]		
[37]	Malignancy of Thyroid Nodules [Age, Gender, Hematocrit, Hemoglobin, Mean Corpuscular Hemoglobin, Mean Corpuscular Hemoglobin Concentration, Mean Corpuscular Volume, Mean Platelet Volume, Platelet Count, Red Blood Cell Count, Red Blood Cell Distribution Width, White Blood Cells, Alkaline Phosphatase, Alanine Transaminase]		
[38]	46 Follicular Thyroid Carcinoma vs. 46 Follicular Adenoma [70 DNA Methylation Haplotype Blocks]		
[39]	Aggressive Extrathyroidal Extension PTC [Magnetic Resonance Imaging Features]		
[40]	Thyroid Peroxidase Activity/Thyroid Hormone Receptor Modulation [Molecular Features from ToxCast Data]		
[41]	Malignancy of Thyroid Nodules [Gray-Level Run-Length Matrix Run-Length Nonuniformity, Maximum Standard Unit Value]		
[42]	Malignancy of Thyroid Nodules [Nodule Size, AP/T ≥ 1 , Solid Component, Micro-Calcifications, Hackly Border, Hypoechogenicity, Presence of Halo, Unclear Border, Irregular Margin, Central Vascularity]		
[43]	Hypothyroidism [Gut Microbiota RNA Features]	Veillonella, Paraprevotella, Neisseria, Rheinheimera	1
[44]	Malignancy of Thyroid Nodules [Size, Shape, Margins, Micro-Calcification, Composition, Echogenicity of the Solid Portion, Halo Sign, Vascularity, Colour Scale Scoring System of Real-Time Elastography, Strain Ratio]	Wreath-Shaped Feature, Micro-Calcification, Strain Ratio	1
[45]	PTC [86 Radiomics Features]		

PTC, papillary thyroid carcinoma.

racy range of logistic regression, the decision tree and the random forest was 84.7–89.7. According to random forest variable importance, tumor diameter, symptoms and extrathyroidal extension were the most important variables for predicting the local recurrence of differentiated thyroid carcinoma. The results of these studies demonstrate that a combination of machine learning and mumerica data is expected to have great utility for predicting the quality of life after thyroid surgery or local recurrence of thyroid cancer.

In a similar vein, a combination of machine learning and genomic data would make great contribution for the early diagnosis of thyroid disease. The aim of a recent study [34] was to use machine learning and genomic data for the early diagnosis of papillary thyroid carcinoma. The source of data was 506 participants enrolled in Cancer Genome Atlas data and their 703 RNA features served as the attributes of this study. Among these attributes, two groups were selected as the most important variables in terms of random forest variable importance for the early diagnosis of papillary thyroid carcinoma: Group 1 including ADD3-AS1 (downregulation), MIR100HG (downregulation), FAM95C (downregulation), MORC2-AS1 (downregulation), LINC00506 (downregulation), ST7-AS1 (downregulation), LOC339059 (downregulation), MIR181A2HG (upregulation), FAM181A-AS1 (downregulation), LBX2-AS1 (upregulation), BLACAT1 (upregulation); Group 2 including hsa-miR-9-5p (downregulation), hsa-miR-146b-3p (upregulation), hsa-miR-199b-5p (downregulation), hsa-miR-4709-3p (upregulation), hsa-miR-34a-5p (upregulation), hsa-miR-214-3p (downregulation). The accuracy of machine learning based on Group 1/Group 2 was decision tree 92.2/87.0, random forest 99.2/99.5, and support vector machine 99.0/98.3.

In a similar context, the purpose of recent research [43] was to adopt machine learning and genomic data for the early diagnosis of hypothyroidism. The sample size of this study was 92 and the attributes of this study were gut microbiota RNA features. Among these features, veillonella, paraprevotella, neisseria and rheinheimera ranked the top in terms of random forest variable importance. Finally, a recent study [44] demonstrates that machine learning together with ultrasound data would provide effective non-invasive decision support systems for predicting the malignancy of thyroid nodules. Data came from 177 thyroid nodules and the following 10 attributes were considered: size, shape, margins, micro-calcification, composition, the echogenicity of the solid portion, halo sign, vascularity, the color scale scoring system of real-time elastography and strain ratio. The random forest showed the best performance in terms of accuracy and the AUC: logistic regression 84.2/92.8, random forest 86.0/93.4, support vector machine 84.8/92.3, gradient boosting 83.7/92.6, and artificial neural network 84.8/90.8. Among the ten attributes, wreath-shaped feature, micro-calcification and strain ratio were the most important variables in terms of random forest variable importance for

predicting the malignancy of thyroid nodules.

4. Discussion

This study reviewed original studies including the random forest and the four other machine learning methods: The twenty original studies were selected out of 33 original studies in PubMed with the search terms “thyroid” (title) and “random forest” (abstract). This study put more focus on the random forest for two reasons. Firstly, it has the advantage of rigorous cross validation from “bootstrap aggregation”: it is a collection of many decision trees with a majority vote on the dependent variable. For example, a random forest with 1000 decision trees samples 1000 training sets with replacements, trains 1000 decision trees with the 1000 training sets, makes 1000 predictions with the 1000 decision trees, and takes a majority vote on the dependent variable. In other words, the random forest with 1000 decision trees uses rigorous 1000-fold cross validation and this explains why it usually shows the best performance together with boosting and neural network approaches [7,15,17,19]. Secondly, the random forest can analyze which predictors are more important for the early diagnosis of a disease [7,15,17,19]. But another method can be more accurate and more appropriate than the random forest in certain circumstances. Little research has been done and more effort is to be made on this topic.

This study reveals that random forest variable importance would vary across different types of data for the early diagnosis of thyroid disease. The following attributes would be important variables in the case of numeric data: (1) clinical stage, marital status, histological type, age, nerve injury symptom, economic income and surgery type for predicting the quality of life 3 months after thyroid cancer surgery; tumor diameter, symptoms and extrathyroidal extension for predicting the local recurrence of differentiated thyroid carcinoma. Likewise, the list of important attributes in the case of genomic data would include: (1) RNA feaures including ADD3-AS1 (downregulation), MIR100HG (downregulation), FAM95C (downregulation), MORC2-AS1 (downregulation), LINC00506 (downregulation), ST7-AS1 (downregulation), LOC339059 (downregulation), MIR181A2HG (upregulation), FAM181A-AS1 (downregulation), LBX2-AS1 (upregulation), BLACAT1 (upregulation), hsa-miR-9-5p (downregulation), hsa-miR-146b-3p (upregulation), hsa-miR-199b-5p (downregulation), hsa-miR-4709-3p (upregulation), hsa-miR-34a-5p (upregulation) and hsa-miR-214-3p (downregulation) for the early diagnosis of papillary thyroid carcinoma; (2) gut microbiota RNA features such as veillonella, paraprevotella, neisseria and rheinheimera for the early diagnosis of hypothyroidism. In a similar vein, the following ultrasound features are expected to request due attention for predicting the malignancy of thyroid nodules: wreath-shaped feature, micro-calcification and strain ratio. As noted before, machine learning is a data-driven method and more

study is to be done for greater external validity. However, the findings above would present useful guidelines on the effective application of random forest variable importance across a variety of data modes for the early diagnosis of thyroid disease in future research.

But current studies on the early diagnosis of thyroid disease based on machine learning has the following limitations. Firstly, many studies adopted cross-sectional data and employing longitudinal data would strengthen the performance of machine learning. Secondly, many studies used data with small sizes in single centers. Using big data (e.g., national health insurance claims data) would make valuable contributions for this area. Thirdly, most studies did not consider possible mediating effects among predictors. Fourthly, some studies reported accuracy or the AUC below 70.0 and these results would not be appropriate as diagnostic tests. Fifthly, binary categories (no, yes) are popular now but they can be refined to multiple categories with more clinical insights. Sixthly, combining different types of machine learning approaches for different types of thyroid data would bring new innovations in many aspects. Finally, it can be noted that this study did not use meta-analysis because different studies would have different diagnostic aims.

5. Conclusions

This article reviewed the recent progress of machine learning for the early diagnosis of thyroid disease. This review demonstrates that machine learning provides an effective, non-invasive decision support system for early diagnosis of thyroid disease.

Abbreviations

AUC, area under the receiver operating characteristic curve.

Author contributions

KSL and HP contributed to research design, data collection, analysis and interpretation, as well as manuscript writing, editing and review. KSL and HP approved the final version of the manuscript.

Ethics approval and consent to participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This research was supported by the Ministry of Science and ICT of South Korea under the Information Technology Research Center support program supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation) (IITP-2018-0-01405).

Conflict of interest

The authors declare no conflict of interest. KSL and HP are serving as the guest editors of this journal. We declare that KSL and HP had no involvement in the peer review of this article and have no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to AP.

References

- [1] Cleveland Clinic. Thyroid disease. 2022. Available at: <https://my.clevelandclinic.org/health/diseases/8541-thyroid-disease> (Accessed: 15 February 2022).
- [2] Mayo Clinic. Thyroid disease: symptoms and treatment. 2019. Available at: <https://www.mayoclinichealthsystem.org/hometown-health/speaking-of-health/thyroid-disease-symptoms-and-treatment> (Accessed: 15 February 2022).
- [3] Thyroid disease—more research needed. *The Lancet*. 2012; 379: 1076.
- [4] Deng Y, Li H, Wang M, Li N, Tian T, Wu Y, *et al*. Global Burden of Thyroid Cancer from 1990 to 2017. *JAMA Network Open*. 2020; 3: e208759.
- [5] Taylor PN, Albrecht D, Scholz A, Gutierrez-Buey G, Lazarus JH, Dayan CM, *et al*. Global epidemiology of hyperthyroidism and hypothyroidism. *Nature Reviews. Endocrinology*. 2018; 14: 301–316.
- [6] Hepp Z, Lage MJ, Espaillat R, Gossain VV. The direct and indirect economic burden of hypothyroidism in the United States: a retrospective claims database study. *Journal of Medical Economics*. 2021; 24: 440–446.
- [7] Lee KS, Ahn KH. Application of artificial intelligence in early diagnosis of spontaneous preterm labor and birth. *Diagnostics*. 2020; 10: 733.
- [8] Lee K, Jung S, Gil Y, Son HS. Atrial fibrillation classification based on convolutional neural networks. *BMC Medical Informatics and Decision Making*. 2019; 19: 206.
- [9] Lee KS, Ahn KH. Artificial neural network analysis of spontaneous preterm labor and birth and its major determinants. *Journal of Korean Medical Science*. 2019; 34: e128.
- [10] Lee KS, Song IS, Kim ES, Ahn KH. Determinants of spontaneous preterm labor and birth including gastroesophageal reflux disease and periodontitis. *Journal of Korean Medical Science*. 2020; 35: e105.
- [11] Lee K, Kim HY, Lee SJ, Kwon SO, Na S, Hwang HS, *et al*. Prediction of newborn's body mass index using nationwide multicenter ultrasound data: a machine-learning study. *BMC Pregnancy and Childbirth*. 2021; 21: 172.
- [12] Lee KS, Kim HI, Kim HY, Cho GJ, Hong SC, Oh MJ, *et al*. Association of preterm birth with depression and particulate matter: machine learning analysis using national health insurance data. *Diagnostics*. 2021; 11: 555.
- [13] Ahn KH, Lee KS, Lee SJ, Kwon SO, Na S, Kim K, *et al*. Predictors of newborn's weight for height: a machine learning study using nationwide multicenter ultrasound data. *Diagnostics*. 2021; 11: 1280.
- [14] Lee KS, Kim ES, Kim DY, Song IS, Ahn KH. Association of gastroesophageal reflux disease with preterm birth: machine learning analysis. *Journal of Korean Medical Science*. 2021; 36: e282.
- [15] Park EK, Lee K, Seo BK, Cho KR, Woo OH, Son GS, *et al*. Machine Learning Approaches to Radiogenomics of Breast Cancer using Low-Dose Perfusion Computed Tomography: Predicting Prognostic Biomarkers and Molecular Subtypes. *Scientific Reports*. 2019; 9: 17847.
- [16] Lee K, Son S, Park S, Kim ES. Automated detection of colorectal

tumors based on artificial intelligence. *BMC Medical Informatics and Decision Making*. 2021; 21: 33.

- [17] Lee JY, Lee K, Seo BK, Cho KR, Woo OH, Song SE, *et al.* Radiomic machine learning for predicting prognostic biomarkers and molecular subtypes of breast cancer using tumor heterogeneity and angiogenesis properties on MRI. *European Radiology*. 2022; 32: 650–660.
- [18] Lee K, Jang J, Yu Y, Heo JS, Han H, Yoon Y, *et al.* Usefulness of artificial intelligence for predicting recurrence following surgery for pancreatic cancer: Retrospective cohort study. *International Journal of Surgery*. 2021; 93: 106050.
- [19] Park HS, Lee KS, Seo BK, Kim ES, Cho KR, Woo OH, *et al.* Machine learning models that integrate tumor texture and perfusion characteristics using low-dose breast computed tomography are promising for predicting histological biomarkers and treatment failure in breast cancer patients. *Cancers*. 2021; 13: 6013.
- [20] Lee K, Park KW. Social Determinants of Association among Diabetes Mellitus, Visual Impairment and Hearing Loss in a Middle-Aged or Old Population: Artificial-Neural-Network Analysis of the Korean Longitudinal Study of Aging (2014-2016) *Geriatrics (Basel, Switzerland)*. 2020; 4
- [21] Lee KS, Park KW. Social determinants of association among cerebrovascular disease, hearing loss and cognitive impairment in a middle-aged or old population: recurrent-neural-network analysis of the Korean Longitudinal Study of Aging (2014–2016). *Geriatrics & Gerontology International*. 2019; 19: 711–716.
- [22] Lee K, Park KW. Artificial Intelligence Approaches to Social Determinants of Cognitive Impairment and its Associated Conditions. *Dementia and Neurocognitive Disorders*. 2020; 19: 114–123.
- [23] Ryu KJ, Yi KW, Kim YJ, Shin JH, Hur JY, Kim T, *et al.* Machine learning approaches to identify factors associated with women's vasomotor symptoms using general hospital data. *Journal of Korean Medical Science*. 2021; 36: e122.
- [24] Lee KS, Kwak HJ, Oh JM, Jha N, Kim YJ, Kim W, *et al.* Automated Detection of TMJ Osteoarthritis Based on Artificial Intelligence. *Journal of Dental Research*. 2020; 99: 1363–1367.
- [25] Lee K, Jha N, Kim Y. Risk factor assessments of temporomandibular disorders via machine learning. *Scientific Reports*. 2021; 11: 19802.
- [26] Song H, Dong C, Zhang X, Wu W, Chen C, Ma B, *et al.* Rapid identification of papillary thyroid carcinoma and papillary microcarcinoma based on serum Raman spectroscopy combined with machine learning models. *Photodiagnosis and Photodynamic Therapy*. 2021; 16: 102647.
- [27] Liu YH, Jin J, Liu YJ. Machine learning-based random forest for predicting decreased quality of life in thyroid cancer patients after thyroidectomy. *Supportive Care in Cancer*. 2021; 30: 2507–2513.
- [28] Kim W, Cho YA, Kim DC, Jo AR, Min KH, Lee KE. Factors associated with thyroid-related adverse events in patients receiving PD-1 or PD-L1 inhibitors using machine learning models. *Cancers*. 2021; 13: 5465.
- [29] Luong G, Idarraga AJ, Hsiao V, Schneider DF. Risk Stratifying Indeterminate Thyroid Nodules with Machine Learning. *Journal of Surgical Research*. 2021; 270: 214–220.
- [30] Yang P, Huang J, Wang Z, Qian L. A predictive model and survival analysis for local recurrence in differentiated thyroid carcinoma. *Minerva Endocrinology*. 2021. (in press)
- [31] Idarraga AJ, Luong G, Hsiao V, Schneider DF. False Negative Rates in Benign Thyroid Nodule Diagnosis: Machine Learning for Detecting Malignancy. *Journal of Surgical Research*. 2021; 268: 562–569.
- [32] Xu S, Huang H, Zhang X, Huang Y, Guan B, Qian J, *et al.* Predictive Value of Serum Thyroglobulin for Structural Recurrence Following Lobectomy for Papillary Thyroid Carcinoma. *Thyroid*. 2021; 31: 1391–1399.
- [33] Hang Y. Thyroid Nodule Classification in Ultrasound Images by Fusion of Conventional Features and Res-GAN Deep Features. *Journal of Healthcare Engineering*. 2021; 2021: 9917538.
- [34] Yang F, Zhang J, Li B, Zhao Z, Liu Y, Zhao Z, *et al.* Identification of Potential lncRNAs and miRNAs as Diagnostic Biomarkers for Papillary Thyroid Carcinoma Based on Machine Learning. *International Journal of Endocrinology*. 2021; 2021: 3984463.
- [35] Qin H, Que Q, Lin P, Li X, Wang X, He Y, *et al.* Magnetic resonance imaging (MRI) radiomics of papillary thyroid cancer (PTC): a comparison of predictive performance of multiple classifiers modeling to identify cervical lymph node metastases before surgery. *Radiologia Medica*. 2021; 126: 1312–1327.
- [36] Sai PV, Rajalakshmi T, Snehalatha U. Non-invasive thyroid detection based on electroglottogram signal using machine learning classifiers. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*. 2021; 235: 1128–1145.
- [37] Olatunji SO, Alotaibi S, Almutairi E, Alrabaa Z, Almajid Y, Altabee R, *et al.* Early diagnosis of thyroid cancer diseases using computational intelligence techniques: a case study of a Saudi Arabian dataset. *Computers in Biology and Medicine*. 2021; 131: 104267.
- [38] Zhang H, Zhang Z, Liu X, Duan H, Xiang T, He Q, *et al.* DNA Methylation Haplotype Block Markers Efficiently Discriminate Follicular Thyroid Carcinoma from Follicular Adenoma. *The Journal of Clinical Endocrinology & Metabolism*. 2021; 106: e1011–e1021.
- [39] He J, Zhang H, Wang X, Sun Z, Ge Y, Wang K, *et al.* A pilot study of radiomics signature based on biparametric MRI for preoperative prediction of extrathyroidal extension in papillary thyroid carcinoma. *Journal of X-Ray Science and Technology*. 2021; 29: 171–183.
- [40] Garcia de Lomana M, Weber AG, Birk B, Landsiedel R, Achenbach J, Schleifer K, *et al.* In Silico Models to Predict the Perturbation of Molecular Initiating Events Related to Thyroid Hormone Homeostasis. *Chemical Research in Toxicology*. 2021; 34: 396–411.
- [41] Aksu A, Karahan Şen NP, Acar E, Çapa Kaya G. Evaluating Focal 18F-FDG Uptake in Thyroid Gland with Radiomics. *Nuclear Medicine and Molecular Imaging*. 2020; 54: 241–248.
- [42] Chen D, Hu J, Zhu M, Tang N, Yang Y, Feng Y. Diagnosis of thyroid nodules for ultrasonographic characteristics indicative of malignancy using random forest. *BioData Mining*. 2020; 13: 14.
- [43] Su X, Zhao Y, Li Y, Ma S, Wang Z. Gut dysbiosis is associated with primary hypothyroidism with interaction on gut-thyroid axis. *Clinical Science*. 2020; 134: 1521–1535.
- [44] Zhao H, Liu J, Lin Q, He Y, Luo H, Peng Y, *et al.* Partially cystic thyroid cancer on conventional and elastographic ultrasound: a retrospective study and a machine learning-assisted system. *Annals of Translational Medicine*. 2020; 8: 495.
- [45] Kwon MR, Shin JH, Park H, Cho H, Hahn SY, Park KW. Radiomics study of thyroid ultrasound for predicting BRAF mutation in papillary thyroid carcinoma: preliminary results. *American Journal of Neuroradiology*. 2020; 41: 700–705.