European
Thyroid Journal

Clinical Thyroidology / Research Article

Eur Thyroid J 2020;9:205–212
DOI: 10.1159/000504882

Received: August 31, 2019
Accepted after revision: November 19, 2019
Published online: December 19, 2019

Machine Learning Prediction of Radiofrequency Thermal Ablation Efficacy: A New Option to Optimize Thyroid Nodule Selection

Roberto Negro[a] Matteo Rucco[b] Annalisa Creanza[c] Alberto Mormile[c] Paolo Piero Limone[c] Roberto Garberoglio[d] Stefano Spiezia[e] Salvatore Monti[f] Christian Cugini[g] Ghassan El Dalati[h] Maurilio Deandrea[c]

[a]Division of Endocrinology, V. Fazzi Hospital, Lecce, Italy; [b]United Technology Research Center, Trento, Italy; [c]Division of Endocrinology and Metabolism, Mauriziano Hospital Umberto I, Turin, Italy; [d]Division of Endocrinology and Metabolism, Molinette Hospital, Turin, Italy; [e]Endocrine Surgery, Ospedale del Mare, ASL NA1 Centro, Naples, Italy; [f]Endocrinology Unit, Azienda Ospedaliera Universitaria Sant'Andrea, Rome, Italy; [g]Radiology Department, Villa Salus Hospital, Venice, Italy; [h]Radiology Department, Policlinico G.B. Rossi, Verona, Italy

Keywords
Thyroid · Nodules · Radiofrequency · Laser · Thermal ablation · Machine learning · Artificial Intelligence

Abstract

Background: Radiofrequency (RF) is a therapeutic modality for reducing the volume of large benign thyroid nodules. If thermal therapies are interpreted as an alternative strategy to surgery, critical issues in their use are represented by the extent of nodule reduction and by the durability of nodule reduction over a long period of time. ***Objective:*** To assess the ability of machine learning to discriminate nodules with volume reduction rate (VRR) < or ≥50% at 12 months following RF treatment. ***Methods:*** A machine learning model was trained with a dataset of 402 cytologically benign thyroid nodules subjected to RF at six Italian Institutions. The model was trained with the following variables: baseline nodule volume, echostructure, macrocalcalcifications, vascularity, and 12-month VRR. ***Results:*** After training, the model could distinguish between nodules having VRR <50% from those having VRR ≥50% in 85% of cases (accuracy: 0.85; 95% confidence interval [CI]: 0.80–0.90; sensitivity: 0.70; 95% CI: 0.62–0.75; specificity: 0.99; 95% CI: 0.98–1.0; positive predictive value: 0.95; 95% CI: 0.92–0.98; negative predictive value: 0.95; 95% CI: 0.92–0.98). ***Conclusions:*** This study demonstrates that a machine learning model can reliably identify those nodules that will have VRR < or ≥50% at 12 months after one RF treatment session. Predicting which nodules will be poor or good responders represents valuable data that may help physicians and patients decide on the best treatment option between thermal ablation and surgery or in predicting if more than one session might be necessary to obtain a significant volume reduction.

Introduction

Thyroid nodules are a common finding in endocrine clinical practice. Their prevalence is approximately 5% when detected by palpation but may occur at 70% frequency when the thyroid is examined by ultrasound [1–3]. Most nodules are benign and asymptomatic. In these

Roberto Negro and Matteo Rucco contributed equally to this work.

karger@karger.com
www.karger.com/etj

Karger

Roberto Negro
Division of Endocrinology, V. Fazzi Hospital
Piazza Muratore
IT–73100 Lecce (Italy)
E-Mail dr.negro@libero.it

cases, they may simply be followed without requiring any treatment [4, 5]. Sometimes, nodules grow over time causing symptoms of compression or esthetic concerns. Until a decade ago, surgery was the only available choice to solve the problem of mass effect in such cases. In recent years, thermal treatments such as radiofrequency (RF) and laser (L) therapy have shown to be a reliable alternative to surgery as a result of their ability to induce a significant reduction of benign nodules that, in most cases, is sustained long term [6–8]. In 5–30% of nodules, two unfavorable outcomes have been observed: a volume reduction rate (VRR) <50% after the first thermal ablation and progressive regrowth [9–13]. These two outcomes represent two sides of the same coin, as it has been shown that nodules, which decrease to a lesser extent, are those which are prone to regrowth in 4–5 years [14]. Data analyses with respect to VRR and its durability of shrinkage should consider baseline nodule volume, ultrasound structure, and the number of procedures required to obtain an optimal reduction in volume [15]. If thermal therapies are interpreted as an alternative strategy to surgery, then critical issues in their use are represented by the VRR, and by the durability of nodule reduction over a long time period. Nodules that recur or for which reduction is unsatisfactory may represent failure when symptoms of compression and esthetic concerns persist following thermal ablation. Thus, predicting the efficacy of thermal treatments may be important in identifying the best candidates for a good response in a single session. Nodules predicted to have a poor response would be addressed surgically or through an additional treatment session. For this purpose, we trained a machine learning model and tested its ability to predict the 12-month VRR of nodules treated with RF.

Methods

The machine learning model was trained with a dataset consisting of 402 Italian patients with cytologically benign nodules subjected to RF at six Italian Institutions (benignity was confirmed by two repeated fine-needle aspirations, Class II of Bethesda reporting system). Data from 337 patients were obtained from a prospective study published by Deandrea et al. [16], and additional data from 65 patients were obtained from the Division of Endocrinology at the Mauriziano Hospital (Turin), and recruited with the same modality. All Institutions shared the same training program, and at least 50 RF procedures had been performed using the same unified protocol and similar devices. The dataset contained patient demographics, ultrasound characteristics of nodules, volume at baseline and 12 months after RF, energy delivered by RF for each nodule, and procedure time.

Table 1. Characteristics of 402 patients with cytologically benign nodules at baseline and 12 months after treatment

Male/female	96/306
Age, years	57.5±12.8
Baseline nodule volume (min–max), mL	25.5±16.9 (3–96)
12-month follow-up nodule volume (min–max), mL	8.9±8.7 (0.2–55)
Echostructure	
Mixed	128 (31.8%)
Spongiform	128 (31.8%)
Solid	146 (36.4%)
Macrocalcifications	
Yes	69 (17.2%)
No	333 (82.8%)
Vascular pattern	
Intense perinodular	139 (34.6%)
Intra- and perinodular	192 (47.8%)
Weak perinodular	71 (17.6%)
Energy delivered, W	60±18
Duration of procedure, min	13±6.5
Energy/mL delivered, J	2,629±1,888
Patients who received mean energy ± 1 SD	346 (86.1%)
Patients who received <mean energy – 1 SD	23 (5.7%)
Patients who received >mean energy + 1 SD	33 (8.2%)

The ultrasound characteristics including echostructure, macrocalcifications, and vascularity were designated as categorical variables (echostructure: #1 mixed, #2 spongiform, #3 solid; macrocalcifications: #0 absent, #1 present; vascularity: #1 intense perinodular, #2 intra- and perinodular, #3 weak perinodular), while nodule volume, energy delivered, time of the procedure and VRR were designated as continuous variables.

We instructed an automatic machine learning model (Operating System: Max OSX Mojave version 10.14.6: Python editor: Spyder version 3.3.1; Python version 3.7; TPOT for automatic machine learning version 0.10.1; scikit-learn version 0.21.2; scipy version 1.2.1) (see online suppl. material; for all online suppl. material, see online Supplementary Materials) by inputting the ultrasound characteristics of 402 selected nodules with the goal of testing the ability of machine learning to predict whether or not a given nodule will exhibit a VRR of at least 50% 12 months after RF treatment [17–20]. We designed and tested an automatic machine learning model using the following variables: (1) baseline nodule volume, (2) echostructure, (3) macrocalcifications, (4) vascularity, and (5) 12-month VRR. Energy delivered and procedure time were not considered input variables because the energy delivered and the time necessary for the procedure were unknown prior to the procedure. The skill of the operator and intraoperative variables, which contribute to the extent of treatment, were not considered input variables.

The "technique efficacy" was conventionally defined as VRR ≥50% [21]. The "primary efficacy rate" was conventionally defined as the percentage of nodules successfully treated following the initial procedure (VRR ≥50%) [21].

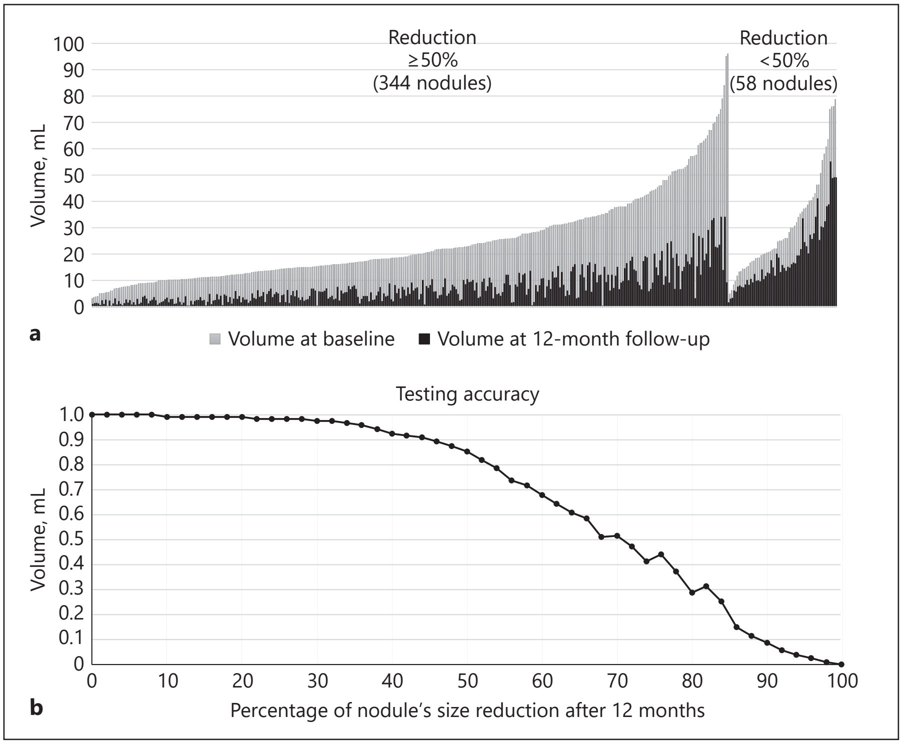

Fig. 1. a Distribution of nodule volume at baseline and at 12 months after treatment, divided by reduction ≥ or <50%. **b** Accuracy of the performances of the trained classifiers on the test set.

Statistical Analyses

Fisher's exact test was used to compare frequencies (percentages) between categorical variables. The Student's *t* test for unpaired data was used to compare continuous variables between two groups. A two-sided *p* value of <0.05 was considered statistically significant. Data were analyzed using the IBM SPSS Statistics version 19 software (SPSS, Chicago, IL, USA). We used the receiver operating characteristic (ROC) curve to show the diagnostic ability of the machine learning model in discriminating nodules with VRR ≥ or <50%. The accuracy corresponds to the area under the curve (AUC) of the ROC. The ROC curve was created by plotting the true positive rate (sensitivity) against the false positive rate (1 – sensitivity). We calculated 95% confidence intervals for accuracy and error. We calculated AUC values, accuracy, sensitivity, and specificity using the Python Scikit-Learn software. We also calculated the likelihood ratios for positive and negative results. We calculated the likelihood ratio for positive results as the sensitivity divided by 1 – specificity and the likelihood ratio for negative results as 1 – sensitivity divided by specificity. The confusion matrix in our study was a 2 × 2 contingency table that reports the number of true positives, false positives, false negatives, and true negatives. The relative weight of the input variables was evaluated by Skater algorithm.

Results

The characteristics of patients and treated nodules are listed in Table 1. At 12 months following treatment, nodules were reduced by 67%. The "primary efficacy rate" was 85.6%. The distribution volume (mL) of thyroid nodules at baseline and 12 months after treatment, divided by VRR ≥ or <50%, is depicted in Figure 1a.

The characteristics of patients and treated nodules with 12-month VRR ≥ or <50% are shown in Table 2. Patients whose nodules exhibited VRR <50% (mean reduction 38%) had significantly larger volume at baseline compared to those with VRR ≥50% (mean reduction 72%). There were no differences in the distribution of echostructure, macrocalcifications, and vascularity. Energy delivered per mL was greater in those with VRR < 50%, whereas the rate of nodules receiving an amount of energy expressed as "mean energy delivered ± 1 SD" was similar between the two groups.

The addition of sex, age, and institution did not affect the results obtained using only echostructure, macrocalcifications, and vascularity as input variables. The ranges representing the highest accuracy of predicting a nodule's reduction were 40–50 and 84–100%. The classifiers in these intervals were extremely accurate in predicting which patients will have VRR <θ, but were not suitable in predicting which patients will have VRR >θ. This was a consequence of the limited number of patients available for study with VRR ≥84%.

For each θ value, we evaluated the performance (accuracy) of the trained classifiers on the test sets (Fig. 1b) and interpreted the graph as follows. For example, for θ = 40,

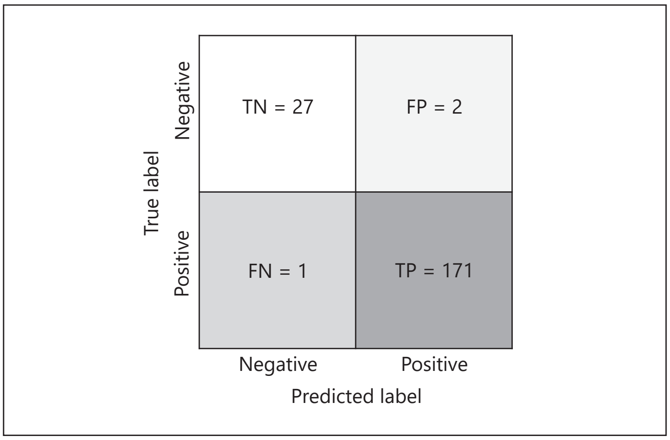

Fig. 2. Confusion matrix. The 2 × 2 contingency table reports the number of true positives, false positives, false negatives, and true negatives; the test set contains 201 patients.

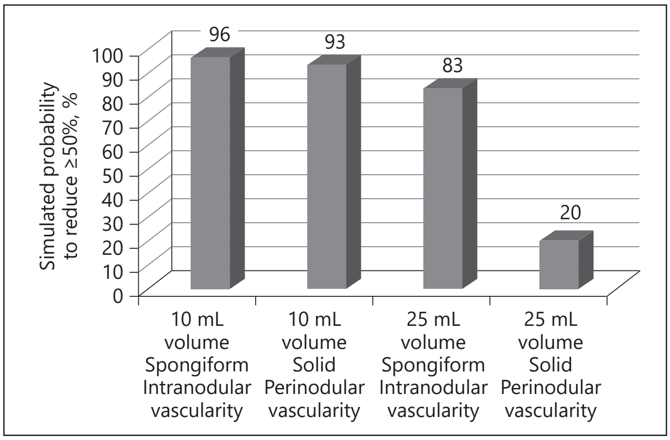

Fig. 3. Simulated probability for a given nodule to reduce ≥50% at 12-month follow-up.

Table 2. Characteristics of patients and treated nodules divided by 12-month reduction ≥ or <50%

Characteristics	12-month reduction		*p*
	≥50%	<50%	
Patients, *n*	344	58	<0.01
Male/female	81/263	13/45	NS
Age, years	58.4±11.8	55.6±10.7	NS
Baseline nodule volume (min–max), mL	24.6±16.4 (3–96)	30.9±18.5 (5–79)	<0.01
12-month follow-up nodule volume (min–max), mL	7.1±6.3 (0.2–34)	19.4±12.5 (2.9–55)	<0.01
Echostructure			
Mixed	111 (32.3%)	17 (29.3%)	NS
Spongiform	109 (31.7%)	19 (32.8%)	NS
Solid	124 (36%)	22 (37.9%)	NS
Macrocalcifications			
Yes	58 (16.9%)	11 (19%)	NS
No	286 (83.1%)	47 (81%)	NS
Vascular pattern			
Intense perinodular	119 (34.6%)	20 (34.5%)	NS
Intra- and perinodular	165 (48%)	27 (46.5%)	NS
Weak perinodular	60 (17.4%)	11 (19%)	NS
Energy delivered, W	59.2±16.3	64.9±25.7	<0.05
Duration of procedure, min	12.8±6.5	14.1±6.7	NS
Energy/mL delivered, J	2,552±1,834	3,084±2,143	<0.05
Patients who received mean energy ± 1 SD	299 (86.9%)	47 (81%)	NS
Patients who received <mean energy – 1 SD	21 (6.1%)	2 (3.5%)	NS
Patients who received >mean energy + 1 SD	24 (7%)	9 (15.5%)	NS

the classifier was able to distinguish between nodules having VRR <40% from those having VRR >40% in 92.5% of cases. For θ = 50, the classifier was able to distinguish between nodules having VRR <50% from those having VRR >50% in 85% of cases. When θ ≥50%, the classification error was 0.15%. There was a 95% likelihood that the confidence interval (0.06, 0.35) covers the true classification error of the model for unseen data. For θ ≥50%, the ROC curve showed an AUC = 0.85; true positive rate = 0.8; false positive rate = 0.2; best cut-off = 1.9.

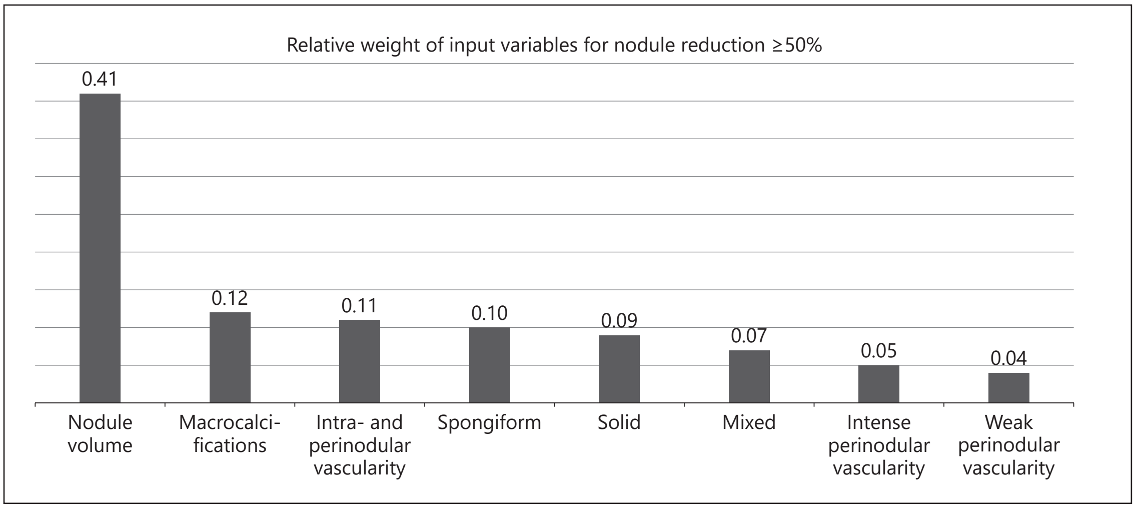

Fig. 4. Input variables sorted by their relative weight.

We also investigated the rate of false positives and false negatives, using the confusion matrix, on the test set containing 201 patients. The classifier correctly classified 171 (true positive) out of 173 nodules having VRR ≥50%, and correctly classified 27 (true negative) out of 28 nodules having VRR <50% (Fig. 2) The performance metrics for the machine learning model, assessed with the validation sets, is shown in Table 3. As an example, we show a simulation resulting from the classifier trained for predicting whether the nodule will have VRR of at least 50% (Fig. 3).

We also investigated the relative weight of the input variables; for the classifier corresponding to θ = 50, the most important factor influencing the 12-month volume reduction was by far baseline volume (Fig. 4).

Discussion

The main finding of this study is that the machine learning model is able, based on ultrasound characteristics, to predict with an elevated degree of accuracy, which nodules undergoing RF will have VRR <50% at 1-year follow-up. The results we obtained warrant implementing a new method to identify which nodules are the best candidates for effective treatment in one single RF session.

Both RF and L are able to induce a significant and durable volume reduction of benign thyroid nodules, thus representing a valid alternative option to surgery [22–26]. When the physician and patient have to decide between a thermal treatment or surgery, it would be important to know how much the nodule will be reduced once treated with thermal ablation and the duration of the effect.

Table 3. Performance metrics for machine learning assessed on the validation sets

	Value	95% CI
Accuracy	0.851	0.802–0.901
Sensitivity	0.698	0.625–0.753
Specificity	0.994	0.983–1.000
Positive predictive value	0.950	0.920–0.980
Negative predictive value	0.952	0.923–0.982

Few studies of RF and L have long-term follow-up data. A Korean multicenter study, consisting of 345 nodules, followed patients for 5 years. The authors reported a significant and durable reduction of nodules and a significant improvement in symptoms and cosmetic scores [27]. It should be noted that few patients reached the 4 or 5-year follow-up milestone (57 and 6, respectively), and two RF sessions were necessary for 20% of the patients, with a mean time to additional RF ablation of 8.5 ± 7.5 months after the initial RF ablation. Another Italian single-center study evaluating 215 patients treated with only one session, found a mean 5-year VRR of 67% [10]. For L, studies with a 3- to 6-year follow-up, showed a reduction rate of approximately 50–60%. Symptoms and cosmetic score decreased significantly in 70–80% of the patients [9, 11–13]. Therefore, studies conducted worldwide demonstrate that both RF and L are effective up to 5 years following thermal treatment.

Analysis of the thermal effects on nodule characteristics indicate that VRR and durability of reduction differ from nodule to nodule. These two outcomes may be influenced by several factors, most importantly baseline volume and energy delivered. Nodule volume exceeding 20–30 mL (>3–4 cm of maximum diameter) and less delivered energy resulted in less efficacy. Literature data indicate that larger nodules receive proportionally less energy as compared to smaller nodules [16, 27].

As the aim of the study is to predict the VRR of each nodule before it is subjected to RF, the energy that will be delivered cannot be established a priori. Thus, the training of the machine learning model cannot take into account the energy that will be delivered to each patient. It has been established by published studies that the energy delivered is a critical factor in determining the final outcome. In our dataset, 86% of the nodules received an amount of energy equal to the mean value ± 1 SD of joules per mL, indicating that there was an elevated degree of homogeneity in the amount of energy delivered per mL in each nodule. Moreover, nodules that were reduced to <50% had a significantly greater volume than those that were reduced to ≥50%. The energy delivered/mL was significantly greater in the former, indicating that nodules that were reduced <50% were not treated with an insufficient amount of energy. In fact, the same percentage of nodules in both groups received an amount of energy equal to the mean value ± 1 SD of joules per mL. There was no difference in the distribution of patients who received the mean –1 SD of energy or the mean +1 SD of energy, suggesting that a similar amount of energy per mL was delivered. Table 2 shows the parity of nodule characteristics (echostructure, calcifications, vascularity); baseline volume represents a critical factor: the larger the nodule, the smaller the chance it will shrink. As RF is an operator-dependent technique, another critical issue pertains to the skill of the operator [28]. At such, it is worth noting that in two studies using RF, the performance of different centers was not significantly different, indicating that the use of shared techniques results in similar VRRs [16, 29]. As a matter of fact, if the performance among different centers was different, the machine learning model would not have been correctly trained and as a consequence, would not have displayed a valid response prediction.

In addition to quantitative measures such as nodule volume and delivered energy, structural features are also crucial. Both with RA and L, spongiform appearance of the nodule and intranodular vascularity are associated with a better response, probably because heat spreads more thoroughly than in compact nodules and periphery-only vascularized nodules [9, 16, 30].

As shown in Figure 3, we performed a simulation of nodules with different volume, structure, and vascularity. A 10-mL nodule with spongiform or solid structure and intranodular or perinodular vascularity had a >90% chance of a >50% reduction. However, a 25-mL nodule had an 83% chance of decreasing >50% if it was spongiform with intranodular vascularity. This chance decreases to 20% if it is solid with weak perinodular vascularity. This simulation helps to understand, for example, that ultrasound characteristics are of little relevance when treating small nodules, but mainly important when treating medium-large nodules. These results were confirmed by the analysis of the relative weight of input variables, which confirmed baseline volume as the major determinant of 12-month VRR.

This information may be relevant when offering a patient thermal treatment. Knowing whether one single session will result in a satisfactory effect or whether more than one session will be required to solve the mass effect is important, not only for the clinical outcome, but also for economic reasons. Few studies have evaluated the cost of thermal treatment versus surgery, and though the costs of one single thermal treatment may appear to be smaller, the cost obviously increases when more than one session is needed [31, 32].

The major limitation to our study is the number of nodules used to train the machine learning model. Though it may not be considered very large, it seemed to be enough to obtain reliable results. The second limitation is the criteria used for nodule classification including echostructure, macrocalcification, and vascularity, which we have autonomously established. The systematic registration of nodule characteristics equally for each patient before recruitment resulted in an increased level of accuracy in collecting data and enabling precise training of the machine learning model. This methodology is mandatory when merging data from different centers, to obtain truthful prediction by the machine learning model.

In conclusion, the present study demonstrates that our machine learning model is reliable and able to identify nodules that will have VRR <50% 12 months after one RFA treatment session. Indeed, as thermal therapies confront surgery, predicting which nodules will be poor responders represents valuable data that may help physicians and patients decide whether thermal ablation or surgery is a better option. Moreover, the data may predict whether more than one session is necessary to obtain a significant volume reduction.

Prospective trials should be performed with prespecified collection of ultrasound characteristics and preestablished energy to be delivered to treat larger number of nodules. This might offer the chance to precisely predict the VRR, and it may specify which nodules will not decrease significantly even following a second thermal procedure.

Statement of Ethics

The research presented in this article was ethically conducted in accordance with the World Medical Association Declaration of Helsinki and the appropriate guidelines for human studies as well as according to animal welfare regulations, including the Animal Research. This study was approved by the institutional review board (IRB). Informed consent from patients was waived by the IRB because of the retrospective nature of this study.

Disclosure Statement

The authors have no conflicts of interest to declare.

Funding Sources

None.

Author Contributions

Roberto Negro and Maurilio Deandrea were responsible for study conception, drafting the manuscript and data evaluation; Matteo Rucco was responsible for data evaluation and statistical analysis; Maurilio Deandrea, Alberto Mormile, Paolo Piero Limone, Roberto Garberoglio, Stefano Spiezia, Salvatore Monti, Christian Cugini, Ghassan El Dalati were responsible for radiofrequency procedures, data collection, and for revising the work critically for important intellectual content.

References

1 Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. Eur J Clin Invest. 2009 Aug;39(8):699–706.

2 Mortensen JD, Woolner LB, Bennett WA. Gross and microscopic findings in clinically normal thyroid glands. J Clin Endocrinol Metab. 1955 Oct;15(10):1270–80.

3 Tan GH, Gharib H. Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. Ann Intern Med. 1997 Feb;126(3):226–31.

4 Hegedüs L, Bonnema SJ, Bennedbaek FN. Management of simple nodular goiter: current status and future perspectives. Endocr Rev. 2003 Feb;24(1):102–32.

5 Durante C, Costante G, Lucisano G, Bruno R, Meringolo D, Paciaroni A, et al. The natural history of benign thyroid nodules. JAMA. 2015 Mar;313(9):926–35.

6 Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, et al.; AACE/ACE/AME Task Force on Thyroid Nodules. American Association of Clinical Endocrinologists, American College of Endocrinology, And Associazione Medici Endocrinologi Medical Guidelines For Clinical Practice for the Diagnosis And Management of Thyroid Nodules—2016 update. Endocr Pract. 2016 May;22(5 Suppl 1):622–39.

7 Kim JH, Baek JH, Lim HK, Ahn HS, Baek SM, Choi YJ, et al.; Guideline Committee for the Korean Society of Thyroid Radiology (KSThR) and Korean Society of Radiology. 2017 Thyroid Radiofrequency Ablation Guideline: Korean Society of Thyroid Radiology. Korean J Radiol. 2018 Jul-Aug;19(4):632–55.

8 Papini E, Pacella CM, Solbiati LA, Achille G, Barbaro D, Bernardi S, et al. Minimally-invasive treatments for benign thyroid nodules: a Delphi-based consensus statement from the Italian minimally-invasive treatments of the thyroid (MITT) group. Int J Hyperthermia. 2019;36(1):376–82.

9 Negro R, Salem TM, Greco G. Laser ablation is more effective for spongiform than solid thyroid nodules. A 4-year retrospective follow-up study. Int J Hyperthermia. 2016 Nov;32(7):822–8.

10 Deandrea M, Trimboli P, Garino F, Mormile A, Magliona G, Ramunni MJ, et al. Long-Term Efficacy of a Single Session of RFA for Benign Thyroid Nodules: A Longitudinal 5-Year Observational Study. J Clin Endocrinol Metab. 2019 Sep;104(9):3751–6.

11 Papini E, Rago T, Gambelunghe G, Valcavi R, Bizzarri G, Vitti P, et al. Long-term efficacy of ultrasound-guided laser ablation for benign solid thyroid nodules. Results of a three-year multicenter prospective randomized trial. J Clin Endocrinol Metab. 2014 Oct;99(10):3653–9.

12 Valcavi R, Riganti F, Bertani A, Formisano D, Pacella CM. Percutaneous laser ablation of cold benign thyroid nodules: a 3-year follow-up study in 122 patients. Thyroid. 2010 Nov;20(11):1253–61.

13 Magri F, Chytiris S, Molteni M, Croce L, Coperchini F, Rotondi M, et al. Laser photocoagulation therapy for thyroid nodules: long-term outcome and predictors of efficacy. J Endocrinol Invest. 2019 Jul. https://doi.org/10.1007/s40618-019-01085-8.

14 Negro R, Greco G. Unfavorable outcomes in solid and spongiform thyroid nodules treated with laser ablation. A 5-year follow-up retrospective study. Endocr Metab Immune Disord Drug Targets. 2019;19(7):1041–5.

15 Huh JY, Baek JH, Choi H, Kim JK, Lee JH. Symptomatic benign thyroid nodules: efficacy of additional radiofrequency ablation treatment session—prospective randomized study. Radiology. 2012 Jun;263(3):909–16.

16 Deandrea M, Garino F, Alberto M, Garberoglio R, Rossetto R, Bonelli N, et al. Radiofrequency ablation for benign thyroid nodules according to different ultrasound features: an Italian multicentre prospective study. Eur J Endocrinol. 2019 Jan;180(1):79–87.

17 Suits DB. Use of dummy variables in regression equations. J Am Stat Assoc. 2012 Apr;52(280):548–51.

18 Olson RS, Urbanowicz RJ, Andrews PC, Lavender NA, Kidd LC, Moore JH. Automating Biomedical Data Science Through Tree-Based Pipeline Optimization. In: Squillero G, Burelli P, editors. Applications of Evolutionary Computation. EvoApplications. Lecture Notes in Computer Science. Volume 9597. Cham: Springer; 2016. pp. 123–37.

19 Goutte C, Gaussier E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In: Losada DE, Fernández-Luna JM, editors. Advances in Information Retrieval. ECIR. Lecture Notes in Computer Science. Volume 3408. Berlin, Heidelberg: Springer; 2005. https://doi.org/10.1007/978-3-540-31865-1_25.

20 Rodríguez JD, Pérez A, Lozano JA. Sensitivity analysis of kappa-fold cross validation in prediction error estimation. IEEE Trans Pattern Anal Mach Intell. 2010 Mar;32(3):569–75.

21 Mauri G, Pacella CM, Papini E, Solbiati L, Goldberg SN, Ahmed M, et al. Image-Guided Thyroid Ablation: Proposal for Standardization of Terminology and Reporting Criteria. Thyroid. 2019 May;29(5):611–8.

22 Trimboli P, Castellana M, Sconfienza LM, Virili C, Pescatori LC, Cesareo R, et al. Efficacy of thermal ablation in benign non-functioning solid thyroid nodule: A systematic review and meta-analysis. Endocrine. 2019 Jul. https://doi.org/10.1007/s12020-019-02019-3.

23 Bernardi S, Dobrinja C, Carere A, Giudici F, Calabrò V, Zanconati F, et al. Patient satisfaction after thyroid RFA versus surgery for benign thyroid nodules: a telephone survey. Int J Hyperthermia. 2018;35(1):150–8.

24 Che Y, Jin S, Shi C, Wang L, Zhang X, Li Y, et al. Treatment of Benign Thyroid Nodules: Comparison of Surgery with Radiofrequency Ablation. AJNR Am J Neuroradiol. 2015 Jul;36(7):1321–5.

25 Sangalli G, Serio G, Zampatti C, Bellotti M, Lomuscio G. Fine needle aspiration cytology of the thyroid: a comparison of 5469 cytological and final histological diagnoses. Cytopathology. 2006 Oct;17(5):245–50.

26 Bellantone R, Lombardi CP, Bossola M, Boscherini M, De Crea C, Alesina P, et al. Total thyroidectomy for management of benign thyroid disease: review of 526 cases. World J Surg. 2002 Dec;26(12):1468–71.

27 Jung SL, Baek JH, Lee JH, Shong YK, Sung JY, Kim KS, et al. Efficacy and Safety of Radiofrequency Ablation for Benign Thyroid Nodules: A Prospective Multicenter Study. Korean J Radiol. 2018 Jan-Feb;19(1):167–74.

28 Pacella CM, Mauri G, Cesareo R, Paqualini V, Cianni R, De Feo P, et al. A comparison of laser with radiofrequency ablation for the treatment of benign thyroid nodules: a propensity score matching analysis. Int J Hyperthermia. 2017 Dec;33(8):911–9.

29 Deandrea M, Sung JY, Limone P, Mormile A, Garino F, Ragazzoni F, et al. Efficacy and Safety of Radiofrequency Ablation Versus Observation for Nonfunctioning Benign Thyroid Nodules: A Randomized Controlled International Collaborative Trial. Thyroid. 2015 Aug;25(8):890–6.

30 Ritz JP, Lehmann KS, Zurbuchen U, Knappe V, Schumann T, Buhr HJ, et al. Ex vivo and in vivo evaluation of laser-induced thermotherapy for nodular thyroid disease. Lasers Surg Med. 2009 Sep;41(7):479–86.

31 Yue WW, Wang SR, Li XL, Xu HX, Lu F, Sun LP, et al. Quality of Life and Cost-Effectiveness of Radiofrequency Ablation versus Open Surgery for Benign Thyroid Nodules: a retrospective cohort study. Sci Rep. 2016 Nov;6(1):37838.

32 Bernardi S, Dobrinja C, Fabris B, Bazzocchi G, Sabato N, Ulcigrai V, et al. Radiofrequency ablation compared to surgery for the treatment of benign thyroid nodules. Int J Endocrinol. 2014;2014:934595.