Significance of FDG–PET standardized uptake values in predicting thyroid disease

Short title: SUVs of thyroid

Tomohiro Kikuchi¹,², Shouhei Hanaoka³, Takahiro Nakao¹, Yukihiro Nomura¹,⁴, Takeharu Yoshikawa¹, MD Ashraful Alam¹, Harushi Mori², Naoto Hayashi¹

1. Department of Computational Diagnostic Radiology and Preventive Medicine, the University of Tokyo Hospital, 7–3–1 Hongo, Bunkyo–ku, Tokyo 113–8655, Japan.
2. Department of Radiology, Jichi Medical University, School of Medicine, 3311–1 Yakushiji, Shimotsuke, Tochigi, 329–0498, Japan.
3. Department of Radiology, the University of Tokyo Hospital, 7–3–1 Hongo, Bunkyo–ku, Tokyo, Japan.
4. Center for Frontier Medical Engineering, Chiba University, 1–33 Yayoicho, Inage–ku, Chiba, Japan.

Corresponding author: Tomohiro Kikuchi
Department of Radiology, Jichi Medical University, School of Medicine, 3311–1 Yakushiji, Shimotsuke, Tochigi, 329–0498, Japan.
Telephone number: +81 285–58–7362
Fax number: +81 285–44–4296
r1419kt@jichi.ac.jp

Keywords: ¹⁸F–fluorodeoxyglucose positron emission tomography, standardized uptake value, thyroid function, thyroid gland, thyroid stimulating hormone

Word count: 3858
Abstract

**Objective:** To determine a standardized cutoff value for abnormal 18F–fluorodeoxyglucose (FDG) accumulation in the thyroid gland.

**Methods:** Herein, 7013 FDG–positron emission tomography (PET)/computed tomography (CT) scans were included. An automatic thyroid segmentation method using two U–nets (2D– and 3D–U–net) was constructed; mean FDG standardized uptake value (SUV), CT value, and volume of the thyroid gland were obtained from each participant. The values were categorized by thyroid function into three groups based on serum thyroid stimulating hormone levels. Thyroid function and mean SUV with increments of 1 were analyzed, and risk for thyroid dysfunction was calculated. Thyroid dysfunction detection ability was examined using a machine learning method (Lightgbm) with age, sex, height, weight, CT value, volume, and mean SUV as explanatory variables.

**Results:** Mean SUV was significantly higher in females with hypothyroidism. Almost 98.9% of participants in the normal group had mean SUV <2 and 93.8% participants with mean SUV <2 had normal thyroid function. The hypothyroidism group had more cases with mean SUV ≥2. The relative risk of having abnormal thyroid function was 4.6 with mean SUV ≥2. The sensitivity and specificity for detecting thyroid dysfunction using Lightgbm were 14.5% and 99%, respectively.

**Conclusions:** Mean SUV ≥2 was strongly associated with abnormal thyroid function in this large cohort, indicating that mean SUV with FDG–PET/CT can be used as a criterion for thyroid evaluation. Preliminarily, this study shows the potential utility of detecting thyroid dysfunction based on imaging findings.
Introduction

Abnormal 18F–fluorodeoxyglucose (FDG) accumulation in the thyroid gland is observed in 5% of patients (1–3). The standardized uptake value (SUV) and pattern of accumulation (focal, diffuse, and diffuse–plus–focal) are used to investigate the clinical significance of abnormal accumulation (2,3). Recently, some studies were performed using radiomics on positron emission tomography (PET) images of the thyroid gland (4,5). Thus, there is continued interest in the significance of abnormal FDG accumulation in the thyroid gland.

Despite this, the SUV in a normal thyroid gland, that should be a key factor for research, has not been sufficiently investigated. Empirically, a normal thyroid gland shows lower FDG accumulation than the blood pool in most cases (6). However, little is known about the actual probability distribution of SUVs. In addition, studies on the appropriate cutoff values for SUV in normal thyroid tissue are lacking. Under these circumstances, it is challenging to ensure that thyroid accumulation is abnormal from an intra– and inter–observer perspective. In addition, research on the role of a high SUV in detecting abnormal thyroid function is lacking.

Therefore, it is important to determine the SUV distribution in the normal population.

The cancer screening program at our institution includes PET/computed tomography (CT) scans for adults over 40 years of age. The data thus collected provides a large repository of thyroid PET/CT scans and same–day information on blood testing. The purpose of this study was to confirm the distribution of SUVs (mean SUV) in the normal population by thyroid function and comprehensively investigate whether mean SUV levels are related to thyroid hormone levels, age, sex, CT value, and thyroid volume. The probability distribution of mean SUV in the thyroid gland and the association between mean SUV and thyroid function was also determined. Additionally, the ability of mean SUV to detect thyroid dysfunction was investigated with and without other factors. This is the first large–scale analysis of the relationship between mean SUVs and thyroid function in healthy individuals.
Methods

Dataset

This study was approved by the ethical review board the University of Tokyo Hospital. Adults who visited our hospital for whole–body medical screening between November 2006 and November 2017 were enrolled. All participants provided written informed consent that their medical images could be used for research purposes. As part of the screening program, PET/CT scans were performed using single–type scanners (Discovery ST Elite, GE Healthcare, Waukesha, WI, USA). Whole–body non–contrast CT images were acquired using the following parameters: field of view (FOV), 500 mm; matrix size, 512×512; voxel size, 0.98×0.98×1.25 mm. The PET images were acquired with the following parameters: FOV, 700 mm; matrix size, 128×128; voxel size, 5.47×5.47×3.25 mm.

Hypothyroidism and hyperthyroidism, defined as elevated and decreased thyroid stimulating hormone (TSH) levels, respectively, are usually diagnosed based on blood test results. These can be defined as subclinical dysfunctions when T4 is within the normal range and overt dysfunctions when T4 is outside the reference values (decreased for hypothyroidism and increased for hyperthyroidism) (7,8). The normal ranges were set to 0.45–4.5 μIU/mL for TSH and 0.8–1.7 ng/dL for T4. In all cases, thyroid hormone levels were determined using blood tests performed on the same day as the PET/CT scan.

Image Processing

We obtained imaging data from our large database by constructing a deep learning–based dedicated automatic segmentation method for the thyroid gland. Our implementation was done using two U–nets. The U–net is a neural network with a symmetric U–shaped architecture with horizontal skip connections that is widely used for semantic segmentation in various fields, including the medical domain (9,10). Depending on the number of dimensions of the image data used for input, it is called a 2D– or 3D–U–net. In general, 3D–U–nets are expected to perform as well as or better than 2D–U–nets (11,12). However, using whole–body CT images as input for 3D–U–net would result in inefficient use of many computational resources. Therefore, we implemented a dedicated automatic segmentation method in two
steps to ensure accurate and efficient segmentation. Step 1 included rough position prediction of the thyroid gland with 2D–U–net and Step 2 included detailed segmentation of the thyroid gland with 3D–U–net (Figure 1).

In Step 1, 2D segmentation was performed using 300 images from the head from whole–body CT images for each case. The slice with the largest area of the thyroid segmentation map was obtained, and the moments (center of gravity from OpenCV library) of the segmentation map in the slice were defined as the center point of the thyroid. Then, CT images were cropped to 96×96×96 voxels centered on the center point. The PET images were resized to fit the voxel size of whole–body CT and cropped at 96×96×96 voxels using the same location information as the thyroid center point.

In Step 2, the 3D–U–net was used for detailed segmentation of the thyroid region: cropped CT images were used as input, and the segmentation map of the thyroid was used as output. Using the segmentation map, the CT value (mean of the segmentation map) and thyroid volume (multiplication of voxel spacing and the number of voxels in the segmentation map) were calculated from CT images, and the mean SUV of the segmented area was calculated using PET images. Before applying this workflow to the entire dataset, 3D–U–net was trained using 107 cases (train 85, validation 22) imaged from January to March 2015. The thyroid ground truth labels, which are binary image data with the same size as the CT image and with the thyroid area set to 1 and the rest of the image set to 0, were created by an experienced radiologist using a web–based image database system (13).

**Statistics and Analysis**

Using 22 cases of validation data, we evaluated whether the entire thyroid region was included in the cropped image from Step 1 and measured the Dice similarity coefficient between the ground truth and the predicted segmentation map in each participant for Step 2. Dice similarity coefficient is a commonly used index for evaluating semantic segmentation results. It can be calculated by the following formula: $2TP / (2TP + FP + FN)$, where TP, FP, and FN are the total number of pixels in the output that are true positives, false positives, and false negatives, respectively, for the ground truth label (14). In addition, the mean SUV, CT value, and volume...
were obtained using both the predicted segmentation map and the ground truth. The differences in each item were calculated, and the 95% confidence intervals (CI) of these differences were evaluated. To evaluate the segmentation results for the entire data set, box–and–whisker plots of thyroid volume were created, and the segmentation results were visually checked for outliers (outside 1.5 times the interquartile range above the upper quartile and below the lower quartile).

The median and quartiles of age, mean SUV, CT values, and thyroid volume were summarized in three groups separately by sex. The differences between the normal and thyroid dysfunction groups were examined using the Mann–Whitney U test, and p < 0.05 was considered significant. The correlation coefficients between age, body height, body weight, CT values, volume, and mean SUV were examined for the normal group.

The mean SUV for the normal group was visualized by creating a histogram and cumulative distribution. The distribution of thyroid function and mean SUV were summarized in a scatter plot and cross–tabulation table. The SUVs were categorized into increments of 1. A chi–square test was used to determine whether the distribution of the hypothyroidism and hyperthyroidism groups differed from that of the normal group. Odds ratios for thyroid dysfunction were calculated with a cutoff value for the mean SUV based on the above cross–tabulation table.

To evaluate the usefulness of SUV in detecting thyroid dysfunction, Lightgbm was used as the classifier, and the performance was evaluated using the area under the curve (AUC). The Youden index was used to determine the appropriate cutoff value. Mean SUV, CT value, thyroid volume, age, sex, body height, and body weight were used for three–class classification: normal, hypothyroidism, and hyperthyroidism. Training and selecting the best model by 10 times cross–validation was done using 70% of the dataset, and the remaining 30% was used to calculate the AUC from the receiver operating characteristic curve to evaluate the performance of the best model. The feature importance of each variable in the trained model was obtained based on “gain” method of Lightgbm library of python3.
Results

Figure 2 depicts the flowchart of inclusion of cases in the study. During the study period, 7773 PET/CT scans were acquired. After excluding patients with incomplete image availability, a history of treatment for thyroid disease, focal abnormalities noted in PET/CT report, and absence of data on thyroid hormone measurements, 7013 cases were included in this study. Among included participants, 182, 265 and 6566 were categorized into the hypothyroidism, hyperthyroidism and normal thyroid level groups, respectively.

The results of the network training with 107 cases are shown in Table 1. In Step 1, the entire thyroid gland was included in the crop area in each case. In Step 2, the median (first and third quartiles) Dice similarity coefficient was 0.84 (0.80, 0.87). The 95% CI of the difference between the predicted segmentation map and ground truth was calculated for the mean SUV, CT value, and volume, all of which included a value of 0 in the range.

Box–and–whisker plots of thyroid volumes calculated from segmentation maps for 7013 cases are shown in Figure 3a. There were no outliers on the lower end of the range, and 144 cases on the upper end. All these cases were visually confirmed by a radiologist, and it was confirmed that there were no cases in which the volume was overestimated due to obvious segmentation errors. The segmentation results of four cases with the largest thyroid volumes are shown in Figure 3b.

The age, mean SUV, CT values, and thyroid volume of the entire dataset are summarized in Table 2. The median age, mean SUV, CT value, and volume in the normal group were 53 years old, 1.07, 81.21 HU, and 16.22 cm$^3$ for males and 55 years old, 1.14, 83.02 HU, and 10.97 cm$^3$ for females. Compared to the normal group, the hypothyroidism group was older and had smaller thyroid volumes in both males (63 years old, 14.63 cm$^3$) and females (60 years old, 9.95 cm$^3$). Additionally, lower CT values in males (78.84 HU) and higher mean SUV (1.32) in females were observed in the hypothyroidism group. In the hyperthyroidism group, CT values and volumes were lower than those in the normal group in both males (75.33 HU, 19.49 cm$^3$) and females (75.24 HU, 13.60 cm$^3$).

Figure 4 shows a histogram of the cumulative distribution curve of the mean SUV in the normal
group. The distribution had a high frequency, between 1 and 1.2, with a long tail.

Figure 5 shows a scatterplot of the mean SUVs and TSH, and Table 3 shows the analysis of thyroid function and mean SUV. In the normal group, 98.9% of participants had mean SUVs of less than 2 and 93.8% of participants with a mean SUV less than 2 had normal function. The chi–square test showed that the hypothyroidism group had more cases with a mean SUV 2 or higher. The odds ratio for thyroid dysfunction was 4.6 (95% CI: 2.8–7.5) for a mean SUV 2 or greater.

Figure 6 shows the correlation between the variables in this study, including the mean SUV in the normal group. None of the factors had a clear correlation with the normal group’s mean SUV and CT value, whereas thyroid volume showed a moderate positive correlation with body height and body weight.

The AUCs for the detection of thyroid dysfunction are shown in Table 4. The AUC of the average was 0.75. The feature importance of the model is shown in Figure 6. In this model, the top three features were image–based information (CT value, Volume, mean SUV).

Discussion

We performed thyroid analysis focusing on SUVs using a large database from a healthy population. This is the first study to analyze mean SUVs on a large scale for normal thyroid function and to examine the utility of SUVs in detecting thyroid function.

Semantic segmentation assigns each pixel in images to a specific class and can provide a map of specific structures, such as organs and lesions (7). We used this method to obtain a segmentation map of the thyroid and automatically calculate imaging information, including SUV, CT value, and volume, resulting in a large–scale analysis. In this study, the mean SUV was used to evaluate FDG accumulation as we considered that maximum SUV might give an erroneous high value owing to segmentation errors. The Dice similarity coefficient between the ground truth and the predicted segmentation map from 3D–U–net was 0.84, sufficient for the following analysis (Table 1). The 95% CI for the differences in mean SUV, CT value, and volume calculated using the predicted segmentation map and ground truth were within a small
range that included 0. In addition, we identified and verified the outliers for thyroid volume to see if there were any significant segmentation errors and found that there were no cases that needed to be actively excluded or corrected (Figure 3). Therefore, we also assumed that there was no significant bias originating from segmentation in the entire data set.

Of the 4648 males and 2365 females included in the analysis, 2.1% of males and 3.7% of females had hypothyroidism, and 4.2% of males and 3.0% of females had hyperthyroidism (Figure 1, Table 3). In community surveys, the prevalence of hypothyroidism and hyperthyroidism varies from 1% to 10% and from 0.5% to 5.9%, respectively (7,15–18). The prevalence of thyroid dysfunction in this study is consistent with previous reports. In the hypothyroidism group, the age distribution was higher, and thyroid volume was smaller in both males and females compared to the normal group (Table 3). In addition, the CT value was lower in males, and the mean SUV was higher in females than in the normal group. These trends are generally consistent with those reported in advanced chronic thyroiditis and atrophic thyroiditis, the most common causes of hypothyroidism (16,17,19,20).

A possible explanation for inconsistency in the statistical significance of imaging findings is that imaging changes and thyroid dysfunction may not occur simultaneously. However, the order of the imaging changes and timeline of thyroid dysfunction remain unclear. For example, in patients with chronic thyroiditis, autoimmune inflammation of the thyroid gland (hyperaccumulation of SUV and/or increase in volume) is expected to precede atrophy (decrease in CT values and/or volume), and a certain percentage of patients develop hypothyroidism. However, this hypothesis has not been entirely proven to date using FDG–PET/CT, and further investigation is required. In the hyperthyroidism group, the CT value was lower, and thyroid volume was larger than those in the normal group, which is consistent with previous reports (21,22). Although a previous report showed increased FDG accumulation in the thyroid gland in Graves' disease, the mean SUV in the hyperthyroidism group in our study was not significantly different from that in the normal group (23). However, as the mean SUV remains unchanged and the volume is larger, total FDG accumulation in the thyroid gland is expected to increase, although the clinical significance is unknown.
In the normal group, the mean SUVs were mostly distributed around 1, and 98.9% of the participants had mean SUVs of less than 2 (Figure 4, Table 2). The distribution for SUVs 2 or higher was more frequent in the hypothyroidism group than that reported in the other groups. This may reflect inflammation caused by chronic thyroiditis; however, this could not be verified in our study due to the absence of a definitive diagnosis. In the hyperthyroidism group, differences in distribution were evidently absent compared with the normal group. If the cutoff value of the mean SUV was set at 2, the odds ratio was 4.6 (95% CI: 2.8–7.5), indicating that a mean SUV of 2 or greater was strongly associated with thyroid dysfunction.

We examined the probability of detection of thyroid dysfunction using the mean SUV with and without other features. When the cutoff for mean SUV was set to 2 (Table 2), the sensitivity was 4.7%, specificity was 98.9%, positive predictive value was 13%, negative predictive value was 93.8%, and the odds ratio was 4.6 (95% CI: 2.8–7.5). This indicated that a mean SUV 2 or higher was strongly associated with thyroid dysfunction. Further, the classification of three groups (normal, hypothyroidism, and hyperthyroidism) using Lightgbm with explanatory variables of mean SUV, CT value, volume, age, sex, body height, and body weight resulted in a macro–average of 0.75. When the cutoff was determined based on the Youden index, the sensitivity was 14.5%, specificity was 99.0%, positive predictive value was 28.6%, and negative predictive value was 97.7%. This was better than the observations stated in Table 2.

Image finding (mean SUV, CT value, and volume) was one of the top three important features for Lightgbm classification in this study (Figure 7). This suggests the possibility to detect abnormal thyroid function to some extent using imaging information and patient background. However, the limited precision of these models restricts clinical application. For future improvements in precision, the use of radiomics features from CT/PET images may be considered. In addition, some biochemical test items, such as liver enzymes and low–density lipoprotein cholesterol, that are known to be associated with thyroid dysfunction, may improve detection accuracy (4,5,24,25). However, it should be noted that the clinical benefit of comprehensive screening for thyroid dysfunction in asymptomatic participants is questionable (26).
Although thyroid dysfunction can be diagnosed easily using blood tests, symptoms can be diverse and nonspecific, leading to a delay in diagnosis in the least suspected cases (27). Our study may provide basis and criteria for suspecting thyroid dysfunction from PET/CT scans, even when these tests are performed for non–thyroid–related reasons. On the other hand, there are many cases of high mean SUVs without thyroid dysfunction or normal mean SUVs with thyroid dysfunction. Whether these groups differ in clinical outcomes from those with high mean SUVs with thyroid dysfunction needs further investigation.

This study had several limitations. First, the impact of selection bias cannot be ruled out. We did not arbitrarily exclude patients. However, all patients were over 40 years of age, that is different from the distribution in the general population. Second, we classified the patients into three groups according to TSH and T4 levels. This indicates that single thyroid disease can be distributed across multiple groups. For example, chronic thyroiditis without abnormal thyroid function was classified as normal. Although we believe that our classification is valid for studying the relationship between SUV levels and thyroid function, the normal group in this study may not be necessarily free of thyroid disease. Third, the effects of artifacts on the thyroid on CT images were not excluded. Thyroid image quality is frequently disturbed by streak artifacts from the bone, such as the scapula and clavicle, that may affect imaging features, especially CT values (28). We performed automatic segmentation and did not set up a process to determine the presence or absence of artifacts in each case. Fourth, the effects of non–thyroidal treatments on the thyroid gland were not considered (e.g., the use of immune checkpoint inhibitors). However, because this study was based on healthy participants who visited our hospital for a whole–body medical screening program, such cases are likely to be rare. Fifth, the effect of small lesions in the thyroid gland was not excluded. Cases in which focal lesions were noted in the PET/CT report were excluded from the analysis, but not in the ultrasonographic findings. Therefore, the analysis includes cases with small lesions that could be depicted only by ultrasonography. Sixth, information on thyroid antibodies was not available in this study. Thyroid antibodies, especially anti–Thyroid Peroxidase (TPO) antibodies, are closely related to chronic thyroiditis, the most common cause of hypothyroidism, and have
also been studied in relation to various indices including CT value, volume, and SUV of the thyroid (21,29,30). If this information could have been included, a more detailed understanding of the pathophysiology might have been possible. In particular, it could be a factor that characterizes false–positive (euthyroid participants with high SUV) and false–negative (hypothyroid participants with low SUV) in the present study.

In conclusion, a cutoff value of 2 for mean FDG SUV can be one crucial indicator to suspect thyroid dysfunction on imaging and be used as a simple assessment of the thyroid gland. The initial results of this study also demonstrate the possibility of detecting thyroid dysfunction based on imaging findings.

Declaration of interest

There is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

This research did not receive any specific grant from any funding agency in the public, commercial or not–for–profit sector.

Author contributions

TK and SH conceived the study. TN and YN significantly contributed to preparing and managing data sets. TY, AA, HM, and NH contributed to the interpretation of the results. All authors reviewed the manuscript draft and critically revised the intellectual content. All authors approved the final version of the manuscript to be published.

Acknowledgements

We would like to acknowledge HIMEDIC Inc. and Siemens Healthcare K.K. that sponsor the Department of Computational Radiology and Preventive Medicine at the University of Tokyo Hospital.
References


20. Maldjian PD & Chen T. Is visual assessment of thyroid attenuation on unenhanced CT of


29. Romaldini JH, Biancalana MM, Figueiredo DI, Farah CS & Mathias PC. Effect of L-thyroxine administration on antithyroid antibody levels, lipid profile, and thyroid volume in patients with Hashimoto’s thyroiditis. Thyroid 1996 6 183–188.

PET/CT in the Management of the Autoimmune Thyroid Diseases. *Frontiers in Endocrinology* 2019 10 208.
**Figure Legends**

**Figure 1. Two steps for getting CT value, Volume, mean SUV of the thyroid.**

Step 1: the 2D–U–net is used. The input was whole–body CT, and the output was the label image. The thyroid "center point" is obtained from the image label, and PET/CT scans were then cropped to 96×96×96 pixels based on the center point.

Step 2: the 3D–U–net was used. The input is a cropped CT image, and the output is a label image. We calculated the CT value and volume from the CT and the mean SUV from the PET image.

**Figure 2. Diagrammatic representation of study inclusion.**

**Figure 3. Box–and–whisker diagram of thyroid volume and segmentation results for four outlier cases.**

(a) Box–and–Whisker diagram created from 7013 thyroid volumes. While there were no outliers on the small side, 144 cases were identified on the large side.

(b) Four cases were selected from the 144 cases in descending order of volume, and the segmentation results were displayed (left: CT image, right: fused CT and mask images). The one with the largest cross–section of the thyroid mask was selected and displayed.

**Figure 4. Histogram with cumulative distribution curve of mean SUV in the normal group.**

The distribution had a single peak and a long tail. The most frequent values ranged from 1 to 1.2, and a few cases exceeded 2.

**Figure 5. Scatterplot of mean SUV and TSH**

*TSH: thyroid stimulating hormone*
Figure 6. Correlation map in normal group.

Correlation coefficients did not exceed 0.2 for mean SUV.

*TSH: thyroid stimulating hormone; T4: thyroxine.

Figure 7. Feature importance of the classification model

The top three features were all derived from PET/CT images.
Table 1. Training results of 2D– and 3D–U–net.

<table>
<thead>
<tr>
<th>Step 1: 2D–U–net</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage of thyroid</td>
<td>100 % (22/22)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 2: 3D–U–net</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dice coefficient</td>
<td>0.84 [0.80, 0.87]</td>
</tr>
<tr>
<td>Difference in CT values</td>
<td>(95% CI) –4.80 to +0.62 HU</td>
</tr>
<tr>
<td>Difference in volume</td>
<td>(95% CI) –0.61 to +1.39 cm³</td>
</tr>
<tr>
<td>Difference in mean SUV</td>
<td>(95% CI) –0.02 to +0.001</td>
</tr>
</tbody>
</table>

CI, confidence interval; CT, computed tomography; SUV, standardized uptake value.
Table 2. Aggregate results of age, mean SUV, CT value, and volume by sex and thyroid function.

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th></th>
<th>Female</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A Normal</td>
<td>B Hypothyroidism</td>
<td>C Hyperthyroidism</td>
<td></td>
</tr>
<tr>
<td>Number of Cases</td>
<td>4354</td>
<td>Subclinical: 80</td>
<td>Subclinical: 185</td>
<td></td>
</tr>
<tr>
<td></td>
<td>53 [46, 62]</td>
<td>63 [57, 74]</td>
<td>51 [45, 59]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.07 (0.97, 1.17)</td>
<td>1.06 (0.94, 1.26)</td>
<td>1.09 (0.99, 1.19)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>81.21 (72.03, 91.09)</td>
<td>78.84 (71.66, 86.97)</td>
<td>75.33 (68.33, 81.44)</td>
<td></td>
</tr>
<tr>
<td>P value†</td>
<td>A:B &lt; 0.001</td>
<td>A:B 0.047</td>
<td>A:B &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A:C 0.057</td>
<td>A:C 0.101</td>
<td>A:B &lt; 0.001</td>
<td></td>
</tr>
</tbody>
</table>

Each cell shows the median and the interquartile range.

†Significant P values < 0.05 are in bold
Table 3. Association of mean SUV and thyroid function.

<table>
<thead>
<tr>
<th>SUV range</th>
<th>&lt;1</th>
<th>1 ≤, &lt;2</th>
<th>2 ≤, &lt;3</th>
<th>3 ≤, &lt;4</th>
<th>4 ≤</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>1654</td>
<td>4842</td>
<td>59</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>38</td>
<td>126</td>
<td>10</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Hyperthyroidism</td>
<td>64</td>
<td>198</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Odds ratio = 4.6 (95% CI: 2.8–7.5) for mean SUV 2 or greater

Significant P values < 0.05 are in bold

SUV, standardized uptake value.
Table 4. AUC values for classification results of thyroid dysfunction detection by Lightgbm.

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal vs Other</td>
<td>0.70</td>
</tr>
<tr>
<td>Hypothyroidism vs Other</td>
<td>0.77</td>
</tr>
<tr>
<td>Hyperthyroidism vs Other</td>
<td>0.78</td>
</tr>
<tr>
<td>Macro–average</td>
<td>0.75</td>
</tr>
</tbody>
</table>

AUC, area under the curve.
Figure 1. Two steps for getting CT value, Volume, mean SUV of the thyroid.

85x51mm (300 x 300 DPI)
Figure 2. Diagrammatic representation of study inclusion.

85x89mm (300 x 300 DPI)
Figure 3. Box-and-whisker diagram of thyroid volume and segmentation results for four outlier cases.
Figure 4. Histogram with cumulative distribution curve of mean SUV in the normal group.
Figure 5. Scatterplot of mean SUV and TSH.

85x84mm (300 x 300 DPI)
Figure 6. Correlation map in normal group.

85x71mm (300 x 300 DPI)
Figure 7. Feature importance of the classification model.

85x50mm (300 x 300 DPI)