Department of Radiology Faculty Papers   Department of Radiology

9-2020

Ultrasonographic risk stratification of indeterminate thyroid nodules; a comparison of an artificial intelligence algorithm with radiologist performance

Aylin Tahmasebi

Shuo Wang

Kelly Daniels

Elizabeth Cottrill

Ji-Bin Liu

See next page for additional authors

Authors

Aylin Tahmasebi, Shuo Wang, Kelly Daniels, Elizabeth Cottrill, Ji-Bin Liu, Jiajun Xu, Andrej Lyshchik, and John R. Eisenbrey

Ultrasonographic risk stratification of indeterminate thyroid nodules; a comparison of an artificial intelligence algorithm with radiologist performance

Aylin Tahmasebi
Radiology
Thomas Jefferson University
Philadelphia, USA
aylin.tahmasebi@jefferson.edu

Shuo Wang
Radiology
Thomas Jefferson University
Philadelphia, USA
Shuo.Wang@jefferson.edu

Kelly Daniels
Otolaryngology
Thomas Jefferson University
Philadelphia, USA
Kelly.Daniels@jefferson.edu

Elizabeth Cottrill
Otolaryngology–Head and Neck Surgery
Thomas Jefferson University
Philadelphia, USA
Elizabeth.Cotrrill@jefferson.edu

Ji-Bin Liu
Radiology
Thomas Jefferson University
Philadelphia, USA
Ji-bin.Liu@jefferson.edu

Jiajun Xu
Ultrasound
Nanjing First Hospital
Nanjing, Jiangsu, China
jj.xujiajun@gmail.com

Andrej Lyshchik
Radiology
Thomas Jeffereson University
Philadelphia, USA
Andrej.Lyshchik@jefferson.edu

John R. Eisenbrey
Radiology
Thomas Jefferson University
Philadelphia, USA
John.eisenbrey@jefferson.edu

Abstract—Background, Motivation and Objective: Thyroid nodules with indeterminate or suspicious cytology are commonly encountered in clinical practice and their clinical management is controversial. Recently, genetical analysis of thyroid fine needle aspiration (FNAs) was implemented at some institutions to differentiate thyroid nodules as high and low risk based on the presence of certain oncogenes commonly associated with aggressive tumor behavior and poor patient outcomes. Our group recently detailed the performance of a machine-learning model based on ultrasonography images of thyroid nodules for the prediction of high and low risk mutations. This study evaluated the performance of a second-generation machine-learning algorithm incorporating both object detection analysis and image classification and subsequently compared performance against blinded radiologists.

Statement of Contribution/Methods: This retrospective study was conducted at Thomas Jefferson University and included an evaluation of 262 thyroid nodules that underwent ultrasound imaging, ultrasound-guided FNA and next-generation sequencing (NGS) or surgical pathology after resection. An object detection and image classification model were employed to first identify the location of nodules and then to assess the malignancy. A Google cloud platform (AutoML Vision; Google LLC) was used for this purpose. Either NGS or surgical pathology was considered as reference standard upon availability. 211 nodules were used for model development and the unused 51 nodules for model testing. Diagnostic performance in 47 nodules for which pathology or NGS were available was compared to blinded reads by 3 radiologists and performance expressed as mean ± standard deviation%.

Results/Discussion: The algorithm achieved positive predictive value (PPV) of 68.31% and sensitivity of 86.81% within the training model. The model was tested on images of 51 unused nodules and all 51 nodules were correctly located (100%). For risk stratification, the model demonstrated a sensitivity of 73.9%, specificity of 70.8%, positive predictive value (PPV) of 70.8%, negative predictive value (NPV) of 73.9% and overall accuracy of 66.7% in the 47 nodules. For comparison, the 3 radiologist performance in this same dataset demonstrated a sensitivity of 53.6±6.6%, specificity of 65.2±6.4%, PPV of 59.7±2.9%, NPV of 59.5±2.1%,and overall accuracy of 59.5±2.2% This work demonstrates that a machine-learning algorithm using image classification performed similarly, if not slightly better than 3 experienced radiologists. Future research will focus on incorporating machine learning findings within radiologist interpretation to potentially improve diagnostic accuracy.

Keywords— Artificial Intelligence, Thyroid Nodule, Ultrasonography, Machine Learning, Next Generation Sequencing

I. INTRODUCTION

In 2020, there are expected to be 52,890 new cases of thyroid cancer and 2180 thyroid cancer related death in the United States alone. This constitutes to 2.9% of all new cancer cases [1]. Through evaluation of incidental findings such as thyroid nodules may have substantial impact on early diagnosis and treatment which ultimately leads to increased survival rate [2]. Ultrasound is widely used as the first line imaging modality for evaluation of thyroid nodules. Presence of high risk or

indeterminate features on ultrasound guide in making decision about subsequent steps in nodule management. Fine needle aspiration (FNA) biopsy followed by cytology to be evaluated using the Bethesda System for Reporting Thyroid Cytopathology has doubled the number of identified thyroid cancers and reduced the number of diagnostic surgical thyroidectomies by half [3]. Most recently Next Generation Sequencing (NGS) significantly improved the risk stratification of indeterminate thyroid nodules by identifying cancer associated genes. Thyroid cancer has specific genetic variations, such as point mutations of proto-oncogenes and chromosomal rearrangements that are related to histopathologic subtype and malignancy [4, 5]. Several studies evaluated different Artificial Intelligence (AI) algorithms for risk stratification of thyroid nodules [6, 7]. In this study we have compared the performance of a Google Auto ML algorithm with three experienced radiologists, taking post surgical pathology or NGS results as the reference standard.

II. METHODS

This retrospective clinical study was approved by the Institutional Review Board (IRB) of Thomas Jefferson University Hospital. Informed consent was waived. Data were retrieved from department Picture Archiving and Communication System (PACS) and consisted of ultrasound images acquired at our institution immediately before or during FNA. Inclusion criteria consisted of all patients who underwent thyroid ultrasound imaging and ultrasound-guided FNA with next-generation sequencing (NGS) with or without surgical pathology between January 2017 and August 2019. An institutional NGS panel of 23 evidence-based gene mutations associated with thyroid malignancy served as a reference to mark FNA samples as high- or low-risk. This 23-gene panel is summarized in Table 1 and served as a rule-in test with samples containing one or more high-risk mutations being classified as high risk for malignancy, whereas samples with no mutation or a mutation considered to be of low or unknown risk were classified as low risk for malignancy by the molecular testing report. In cases where total thyroidectomy or lobectomy were performed following ultrasound imaging, subsequent malignant or benign pathology were treated as high or low risk respectively for the purpose of this study.

Table 1: High Risk Genes on NGS Used as a Reference Standard

*aa denotes amino acid residue numbers.

Gene	Human Genome Region
AKT1	aa 17-18
APC	aa 178-291 and 312-1594 with splice sites
AXIN1	aa 1-688 and 731-865 with splice sites
BRAF	aa 594-606, 439-478
CDKN2A	Full with splice sites
CTNNB1	aa 6-60
DNMT3A	aa 881-883
EGFR	Exons 18,19,20,21
EIF1AX	aa 1-6, 35-86, and 115-147
GNAS	aa 201-203 and 226-227
HRAS	aa 10-14 and 60-62 and 146
IDH1	aa 67-71, 123-134
KRAS	aa 10-14 and 60-62 and 146
NDUFA13	Full with splice sites
NRAS	aa 10-14 and 60-62 and 146
PIK3CA	aa 520-554 and and 980-1069
PTEN	Full with splice sites
RET	Aa 883, 918, 588-636
SMAD4	aa 36-552 with splice sites
TERT	Promoter chr5:1295228 and 1295250
TP53	aa 26-393 with splice sites
TSHR	Full with splice sites
VHL	Full with splice sites

262 thyroid nodules that underwent ultrasound imaging, ultrasound-guided FNA and next-generation sequencing (NGS) or surgical pathology after resection were used. patient information, manufacturer label, and scale bars were removed via a cropping script written in Matlab (2016a, The Mathworks Inc., Natick, MA) (Fig. 1). After the de-identification, approximately 80% of cases (training set) were uploaded into the Google AutoML for object detection and image classification running on the cloud platform. An experienced radiologist who was blinded to the NGS and pathology results used bounding boxes and labels to mark the location and the type of the lesion as well as an area that the lesion was included. Once the training was completed, 20% of cases (prediction set) which were model-naïve were uploaded to the pre-trained deployed model to evaluate the model performance. Three experienced radiologists classified lesions as high or low risk based on the presence of very hypoechoic, taller-than-wide, extra-thyroidal extension and punctate echogenic foci features of ultrasound images. All reads and predictions were compared to NGS or pathology (when available) as a reference standard. Statistical analysis was performed in GraphPad Prism Version 8.4.2 for Windows, GraphPad Software, La Jolla California USA. Data was presented as mean ± standard deviation. Diagnostic performance in 47 nodules out of 51 for which pathology or NGS were available was compared to blinded reads by 3 radiologists.

III. RESULTS

The internal validation within the object detection model resulted in positive predictive value (PPV) of 68.31%, and a sensitivity of 86.81%. When this model was applied to the 51

prediction images it correctly identified the location of nodule in all 51 (100%) cases.

For risk stratification, the model demonstrated a sensitivity of 73.9%, specificity of 70.8%, positive predictive value (PPV) of 70.8%, negative predictive value (NPV) of 73.9% and overall accuracy of 66.7% in the 47 nodules. For comparison, the 3 radiologist performance in this same dataset demonstrated a sensitivity of $53.6 \pm 6.6\%$, specificity of $65.2 \pm 6.4\%$, PPV of $59.7 \pm 2.9\%$, NPV of $59.5 \pm 2.1\%$, and overall accuracy of $59.5 \pm 2.2\%$ (Fig. 2).

IV. CONCLUSION

AI demonstrated that the algorithm using image classification performed similarly, if not slightly better than 3 experienced radiologists. The object detection model correctly identified 100% of lesions. AI has lots of potentials to be implemented in daily radiology practice as an auxiliary tool to improve workflow and as a second opinion for more accurately identifying cases which need further evaluations to reduce unnecessary invasive procedure. Future research will focus on incorporating machine learning findings within radiologist interpretation to potentially improve diagnostic accuracy.
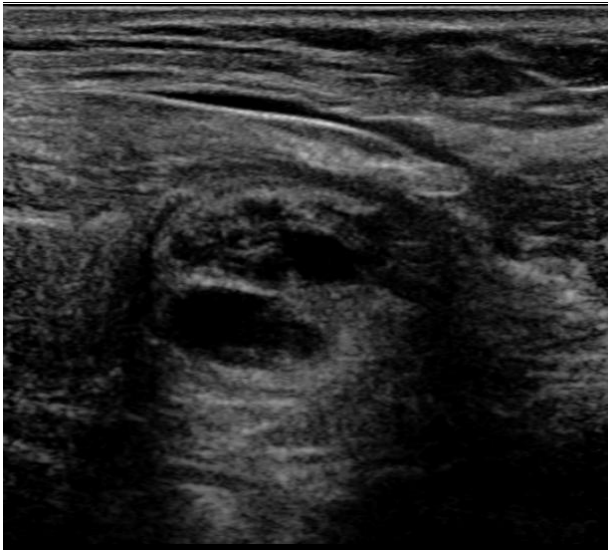
REFERENCES

[1] National Cancer Institute, Surveillance, Epidemiology and End Results Program. Cancer Stat Facts: Thyroid Cancer. Available at www.seer.cancer.gov/statfacts/html/thyro.html. Accessed September 6, 2020.

[2] Kamran, S.C., et al., *Thyroid nodule size and prediction of cancer.* J Clin Endocrinol Metab, 2013. **98**(2): p. 564-70.

[3] Gharib, H. and J.R. Goellner, *Fine-needle aspiration biopsy of the thyroid: an appraisal.* Ann Intern Med, 1993. **118**(4): p. 282-9.

[4] Daniels, K.E., et al., Diagnostic Value of TI-RADS Classification System and Next Generation Genetic Sequencing in Indeterminate Thyroid Nodules. Acad Radiol, 2020.

[5] Nishino, M., Molecular cytopathology for thyroid nodules: A review of methodology and test performance. Cancer Cytopathol, 2016. **124**(1): p. 14-27.

[6] Ko, S.Y., et al., Deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. Head Neck, 2019. **41**(4): p. 885-891.

[7] Daniels, K., et al., *Machine Learning by Ultrasonography for Genetic Risk Stratification of Thyroid Nodules.* JAMA Otolaryngol Head Neck Surg, 2019: p. 1-6.

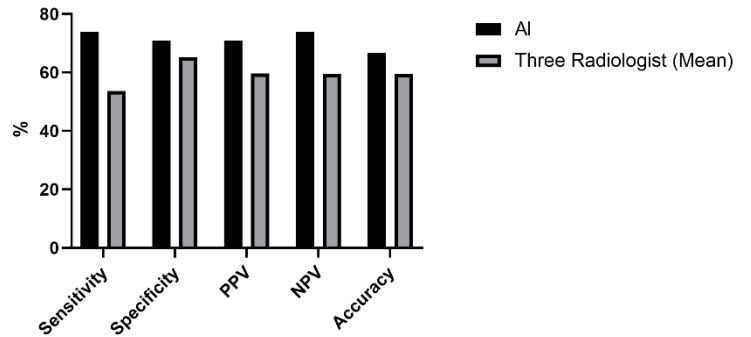Figure 1. Example of a deidentified image fed into Google AI

Figure 2. Risk stratification comparison between AI algorithm and the three experienced radiologists.