Machine Learning-Based Ultrasound Radiomics for Predicting Distant Metastasis in Follicular Thyroid Carcinoma: Performance Evaluation and Clinical Implications

Ji-hoon Kim MD, Department of Radiology, Seoul National University Hospital, Seoul, Korea

Introduction

Background: FTC (10–15% of differentiated thyroid cancers) is notorious for hematogenous distant metastasis — a key prognostic determinant. Unlike PTC, FTC rarely shows suspicious US features, limiting preoperative risk stratification. K-TIRADS categorisation has demonstrated limited utility in identifying high-risk FTC. ML-based radiomics have emerged as promising analytical tools, yet remain unvalidated for rare oncologic outcomes. TERT promoter and RAS mutations are associated with aggressive behavior, but integrative risk models are lacking.

Objective: To evaluate ML-based US radiomics performance for predicting FTC distant metastasis, together with various clinical, imaging (K-TIRADS), and molecular risk factors.

Materials and Methods

Study Design & Patients
- Retrospective single-centre cohort; 105 surgically confirmed FTC patients
- 8 with distant metastasis / 97 without; median follow-up 10.6 yrs (IQR 9.4–17.1)
- Inclusion: thyroid lobectomy/total thyroidectomy + FFPE tissue available for TERT/RAS analysis

Molecular Analysis
- Sanger sequencing of KRAS, HRAS, NRAS and TERT promoter from FFPE samples
- Each sample sequenced twice; reviewed by expert thyroid pathologist

US Imaging & Radiomics
- Preoperative US reviewed by 2 radiologists (16 & 7 yrs exp.); K-TIRADS classification applied
- 106 radiomics features via PyRadiomics (3D Slicer): shape, first-order, GLCM, GLDM, GLRLM, GLSZM, NGTDM
- ROI manually segmented (consensus, ≥5 yrs experience)

ML Classifiers & Validation
- Five classifiers: SVM, Logistic Regression, Naïve Bayes, Random Forest, KNN
- 3-fold cross-validation + mRMR feature selection; AUC, accuracy, sensitivity, specificity assessed
- Benchmarked vs. age, tumour volume, and TERT–RAS co-mutation status

Statistical Analysis
- Categorical: χ^2/Fisher exact; Continuous: ANOVA/Student t-test (R software)
- Multivariable: Firth's penalised logistic regression; p<0.20 entry threshold
- ML pipelines: scikit-learn (Python); significance p<0.05

Results

Clinicopathologic Summary
- 105 FTC patients; F:M = 84:21; mean age 45.1±14.2 yrs
- All 8 metastatic cases: minimally invasive FTC; Bethesda II most frequent (57.1%)
- 4/8 (50%) metastatic cases had benign preoperative cytology (Bethesda II)
- Metastasis diagnosed at median 9.5 yrs post-surgery (range 0–16 yrs); 2 concurrent at surgery

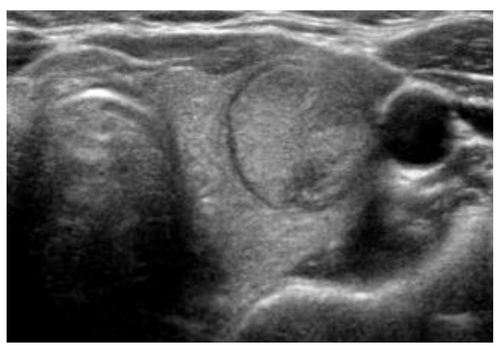

Fig. 1. **Non-metastatic FTC (F/28, no mutation)**
Solid isoechoic nodule, parallel orientation, smooth margin (left thyroid). No mutation identified.

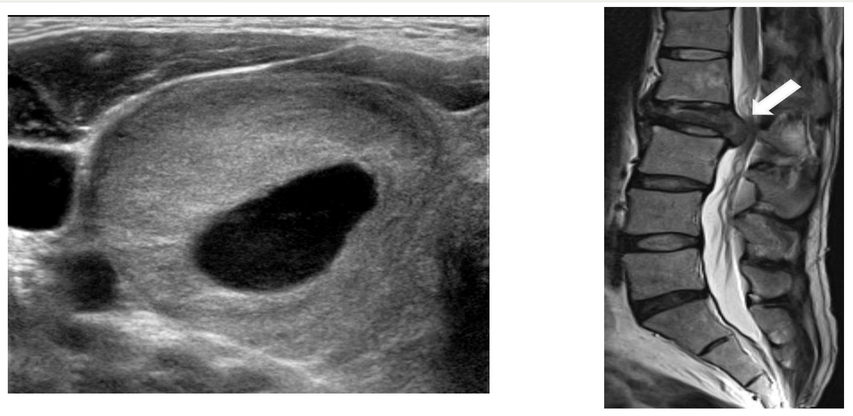

Fig. 2. **M/37, TERT–RAS co-mutation, spine metastasis**
A. US: predominantly solid isoechoic nodule, smooth margin, right thyroid. B. T2 sagittal MRI: pathological fracture L2 (arrow) from FTC metastasis.

Table. ML Radiomics Classification (mRMR Feature Selection, Test Set

Classifier	Accuracy	Sensitivity	Specificity	AUC
SVM	0.9048	0.9307	0.2500	0.5682
Logistic Reg.	0.8952	0.9300	0.2000	0.5631
Naive Bayes	0.8190	0.9239	0.0769	0.5223
Random Forest	0.9143	0.9231	0.0000	0.5000
KNN	0.9143	0.9231	0.0000	0.4948
Average	**0.8895**	**0.9262**	**0.1054**	**0.5297**

AUC = area under ROC curve; KNN = k-nearest neighbor; mRMR = minimal redundancy maximal relevance; Spec = specificity

Key Results
- TERT–RAS co-mutation: 75.0% in metastatic vs. 3.1% in non-metastatic group (p<0.001)
- Diagnostic performance of TERT–RAS co-mutation: Sensitivity 75%, Specificity 96.9%, NPV 97.9%
- Metastatic group significantly older (61.1±12.2 vs. 43.8±13.5 yrs, p=0.001) and larger tumour volume (44.7±42.7 vs. 15.9±29.8 cm³, p=0.013)
- US features (echogenicity, margin, K-TIRADS): no significant differences between groups
- ML radiomics AUC on test set: mean 0.53 (range 0.49–0.57) — near chance level; poor specificity (mean 10.5%)
- All 8 metastatic cases were minimally invasive FTC; metastatic sites: lung (5), rib (2), spine (1), iliac bone (1)

Table. Multivariable Analysis: Independent Predictors of Distant Metastasis (Firth's Penalised LR)

Variable	OR	95% CI	p-value
Age (per year)	1.08	1.00–1.20	0.040*
Tumour volume (cm³)	0.99	0.91–1.05	0.783
Maximal diameter (cm)	1.68	0.53–8.26	0.384
TERT–RAS co-mutation	**52.1**	**7.21–822.5**	**<0.001***

*p<0.05. OR = odds ratio; CI = confidence interval. Variables with p<0.20 in univariable analysis entered into model.

Conclusions

Current Findings
In this real-world dataset, ML-based US radiomics failed to predict FTC distant metastasis meaningfully (test AUC 0.53), reflecting fundamental limitations when applied to rare outcomes with severe class imbalance (8 vs. 97). TERT–RAS co-mutation and age remain the most clinically actionable independent predictors.

Future Directions for AI Integration
- **Multimodal AI:** Integrate US radiomics with CT/MRI features, molecular markers (TERT/RAS), and clinical variables into a unified model — transcending single-modality limitations
- **Larger Multi-centre Datasets:** Rare oncologic events require international consortia; federated learning enables privacy-preserving data sharing across institutions
- **Class Imbalance Strategies:** Synthetic oversampling (SMOTE, GAN-based augmentation) and cost-sensitive learning to address the rarity of distant metastasis events
- **Foundation Models & Transfer Learning:** Large-scale pretrained vision models (e.g. SAM, DINO) fine-tuned on thyroid US may extract richer features than hand-crafted radiomics
- **Explainable AI (XAI) & Prospective Validation:** Transparent, interpretable models with prospective design and preoperative biopsy-integrated molecular data are essential before clinical deployment