

ASSESMEMENT OF AN ARTIFICIAL INTELLIGENCE-BASED DECISION SUPPORT SYSTEM IN THE THYROID NODULE EVALUATION IN CLINICAL PRACTICE.

Pablo Fernández Velasco, Beatriz Torres Torres, Paloma Pérez López, Katia Benito-Sendín Plaar, Daniel de Luis, Gonzalo Díaz Soto

Objective To evaluate the impact of a decision support system (DSS) based on artificial intelligence (AI) -KoiosDS- on ultrasound (US) image analysis and risk stratification in thyroid nodules.

Material and methods Retrospective US study of all thyroid nodules with histologic (AP) result from June 2020 to December 2021. Diagnostic performance of US with ACR-TIRADS, by four endocrinologists and DSS, before and after the use of AI was evaluated.

Results A total of 172 patients (83.1% women) with a mean age of 52.3 ± 15.3 years were evaluated. The maximum nodular diameter was 2.9 ± 1.2 cm, with 10.7% being malignant. 81.4% and 24.5% of the nodules classified by DSS as ACR-TIRADS 3 and 4, respectively, were reclassified into lower risk categories with AI

When performing a ROC curve to assess the diagnostic performance of endocrinologists and DSS against the AP diagnosis, a mean increase in the area under the curve (AUC) after the use of AI was observed for endocrinologists (AUC=0.730 vs. 0.790, $p < 0.001$) and IP (AUC=0.696 vs. 0.735, $p < 0.001$).

When evaluating the impact of AI on diagnostic accuracy, we observed an improvement in mean sensitivity (S) (82.75% vs. 88%), specificity (S) (35.25% vs. 50.25%), a high negative predictive value (NPV) (94.5% vs. 96.75%) and an increase in positive predictive value (PPV) (13.5% vs. 17.75%).

When analyzing the degree of agreement in the US characteristics, we observed a mean increase in concordance with the use of AI in all ultrasound patterns, especially echogenicity (κ 0.456 vs. 0.642; $p < 0.001$). As well as an improvement in the mean correlation between the endocrinologist's ACR TIRADS scores after the use of AI ($r=0.678$ vs. 0.801, $p < 0.001$).

When estimating the DSS as a learning tool for the endocrinologists in the first 25 image evaluated by the endocrinologist compared to the last 25, an improvement in the degree of agreement after using the AI was observed in the assessment of all the ultrasound features, particularly echogenicity (κ 0.474 ± 0.244 vs 0.698 ± 0.111 ; $p < 0.001$).

Conclusion The use of AI in an Endocrinology Department was associated with an overall improvement in the diagnostic ability of US, as well as an increase in S, E, NPV and PPV. AI reclassified more than half of the nodules with intermediate ACR TIRADS into lower risk categories. All US ACR-TIRADS patterns increased the degree of agreement and interindividual variability were reduced with the use of AI. AI proved to be a useful learning tool in the evaluation of thyroid nodule.