

Background

The Ovarian Diagnostic Score (ODS) is an end-to end model integrating Deep Learning for segmentation and Radiomics for classification developed on a dataset from Imperial College Healthcare NHS Trust¹. The ODS model was externally validated on two independent test sets from Forli (Italy) and Seoul (South Korea) and reached an AUC of 0.90.

Objectives

The aim of this study was to determine the performance of the ODS model in a real-world adnexal mass dataset.

Methods

- The ODS model combines automated segmentation using a Multi-Attention Network with radiomic feature extraction and classification using ridge regression.
- The external validation dataset consisted of women with an adnexal mass on ultrasound who underwent surgery in a University Hospital in Ukraine (UKR) (n=244, 338 images).
- The reference standard was histological diagnosis.
- Compressed images were converted into TIFF format for segmentation, then NIfTI for radiomics feature extraction using TexLAB 2.0.
- All classifiers co-developed alongside the ODS model were evaluated, including Random Forest and XGBoost.
- The area under the ROC curve (AUC), sensitivity and specificity were used to measure classification performance.

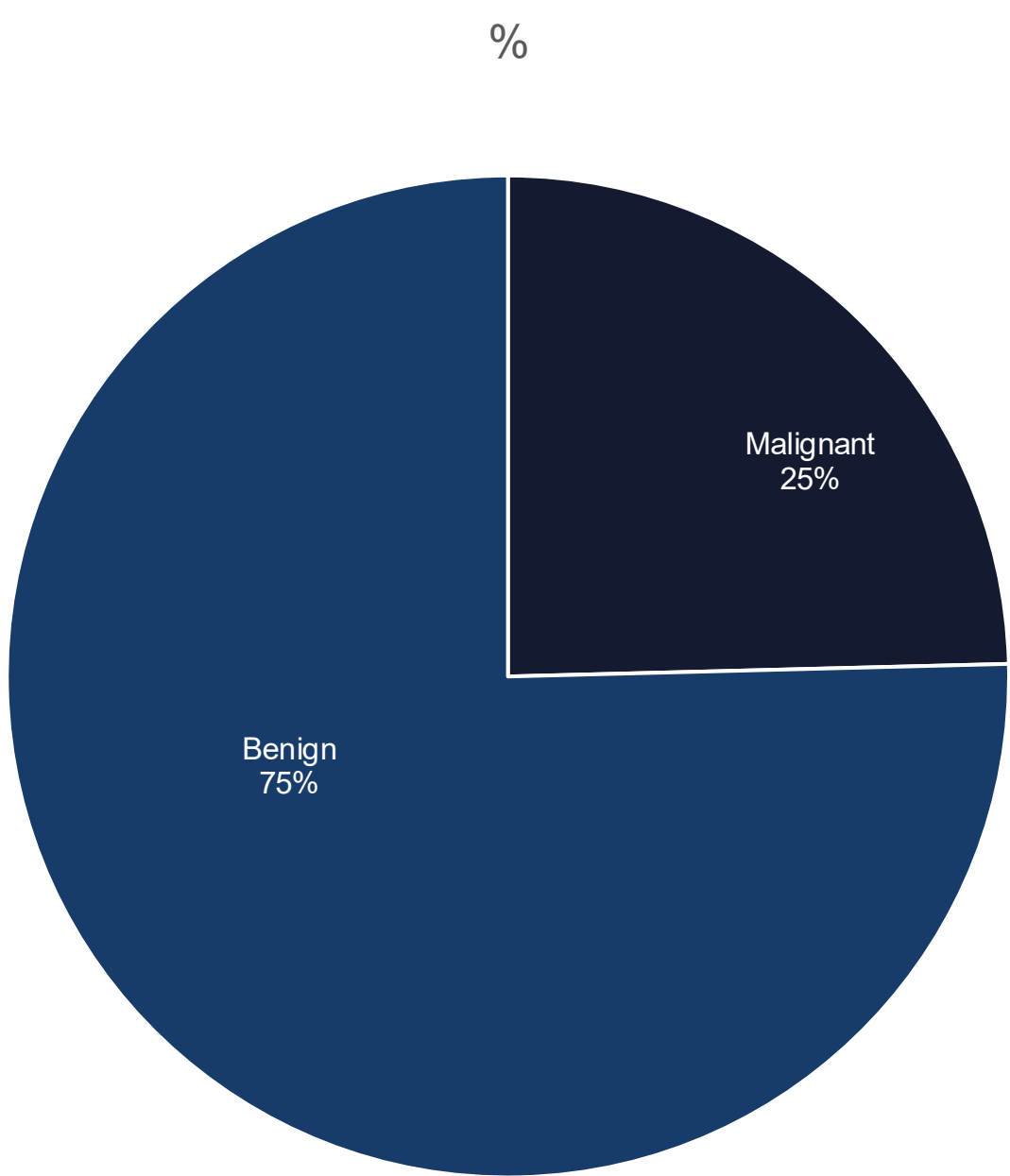


Figure 1: The histological classification of the Ukraine adnexal mass dataset.

Machine Learning technique	Area under the ROC curve	Sensitivity	Specificity
ODS Model	0.60	0.84	0.30
Random Forest combined with isotonic Youden J threshold optimisation	0.72	0.71	0.64

Table 2: Performance of ODS model and the best machine learning techniques at the classification of the Ukraine adnexal mass dataset

	Imperial n=577	Ukraine n=244
Image size	1136 x 852 pixels	1608 x 1361 pixels
Image type	TIFF (ultrasound system)	TIFF (converted)
Image quality	Raw	Compressed
Calipers	0%	16%
Grayscale	100%	76.1%
Transvaginal ultrasound approach	100%	91.0%

Table 1: Comparison of Imperial College Healthcare NHS Trust (ODS model) and Ukraine adnexal mass datasets

Conclusion

- Image type, size and processing may affect radiomics feature patterns, based on a drop in ODS model performance in this compressed image dataset compared to previous external validation (AUC 0.90).
- Random Forest appears to perform marginally better than the ODS model in this dataset.
- Standardised image acquisition protocols are key to the development of robust machine learning models.

1.Barcroft JF, Linton-Reid K, Landolfo C, Al-Memar M, Parker N, Kyriacou C, Munaretto M, Fantauzzi M, Cooper N, Yazbek J, Bharwani N, Lee SR, Kim JH, Timmerman D, Pasma J, Savelli L, Saso S, Aboagye EO, Bourne T. Machine learning and radiomics for segmentation and classification of adnexal masses on ultrasound. NPJ Precis Oncol. 2024 Feb 20;8(1):41. doi: 10.1038/s41698-024-00527-8.