Device Impact on Machine Learning Classifier Accuracy in Detecting Cervical Dysplasia

TAP TO RETURN TO KIOSK MENU

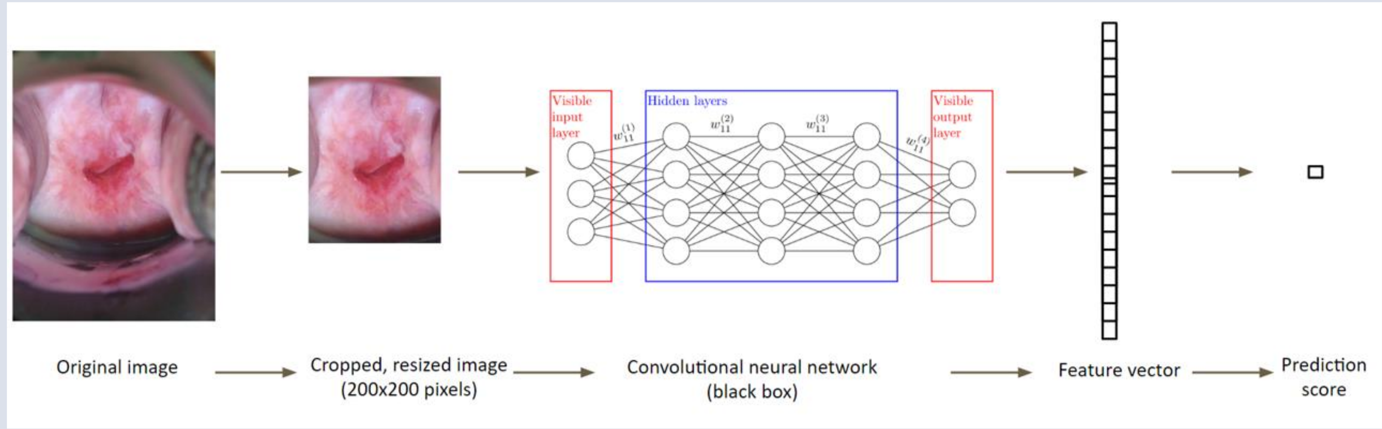
KC Fernandes[1], T Freitas[1], Y Zall[2], R Nissim[2], D Levitz[2]

1 NILG.ai, Porto, Portugal; 2 MobileODT Ltd., Tel Aviv, Israel

Introduction

Automated visual evaluation (AVE) is a promising technology that uses a machine learning (ML) classifier to predict the likelihood of pathology in a cervical image. AVE is accurate, fast, and inexpensive, and thus it has tremendous potential.

Because AVE is based on ML and not an in vitro assay, very little is known about which features affect its performance, and which do not.

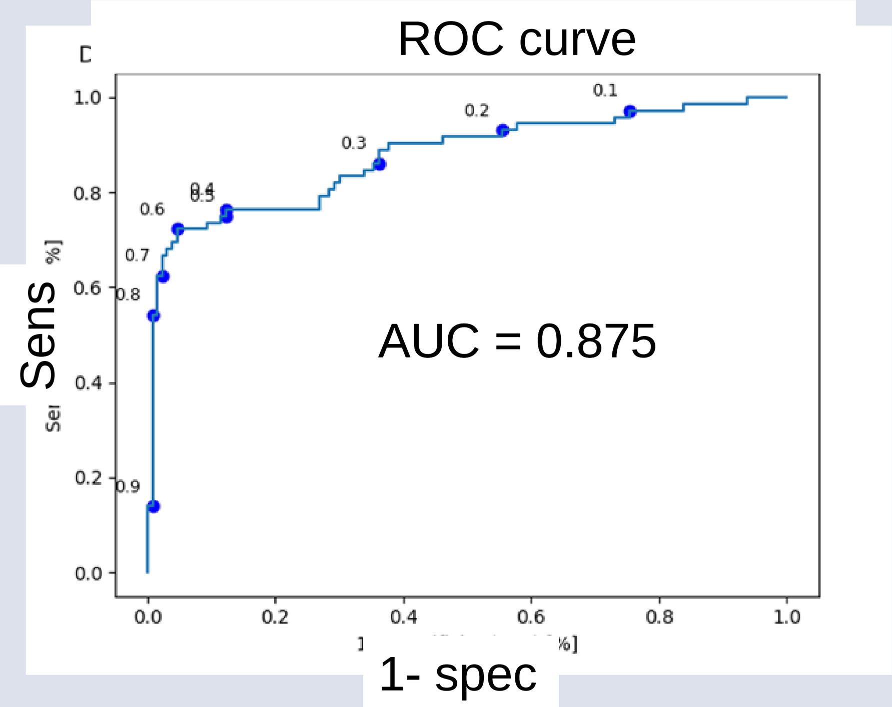


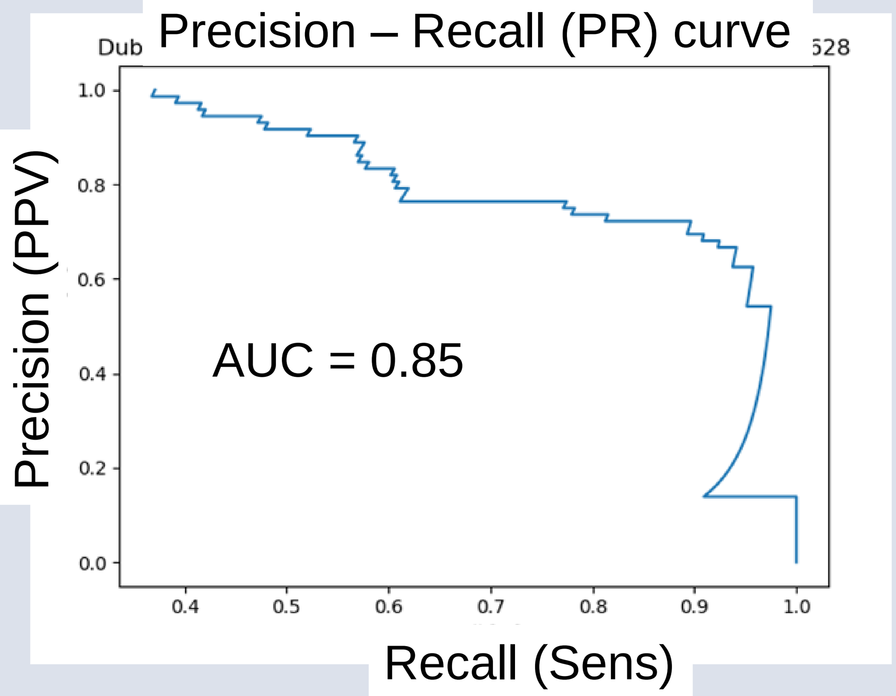
This analysis examines the performance of AVE under different imaging conditions. There are 2 types of analyses presented. In one type of analysis, the images are modified by a specific effect through image processing and manipulation, and the performance drop off is measured. The other analysis involves grouping image metadata and comparing different sub-cohorts.

The **goal** of this study is to characterize which imaging features AVE is sensitive to. The analysis was done on a global set of images collected by users of the Enhanced Visual Assessment (EVA) System across the globe.

Methods

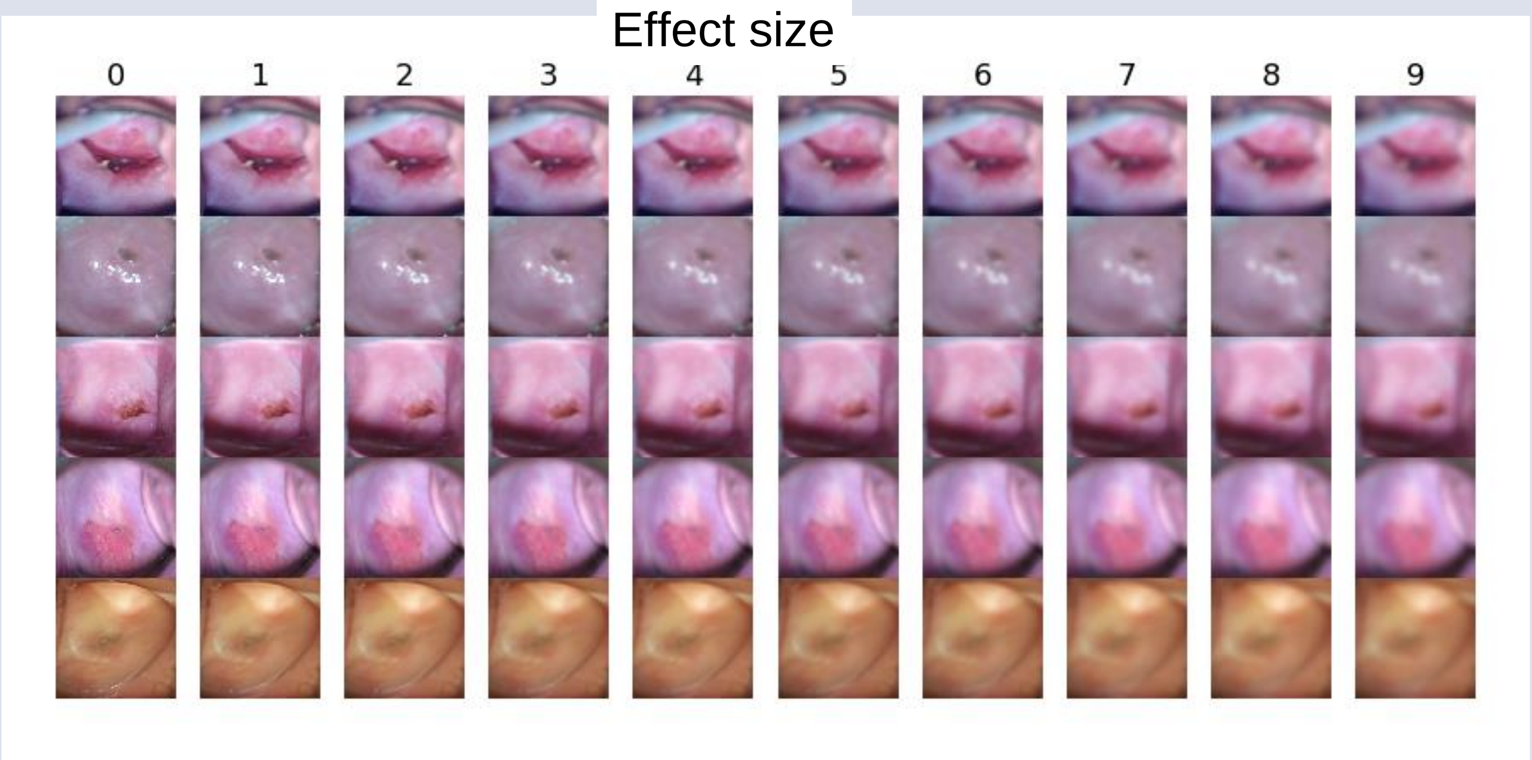
A data set comprising of images from N=202 patients (72 abnormal, 130 normal) was used for testing an AVE classifier. The global distribution of data, as well as base classifier performance, are shown below.

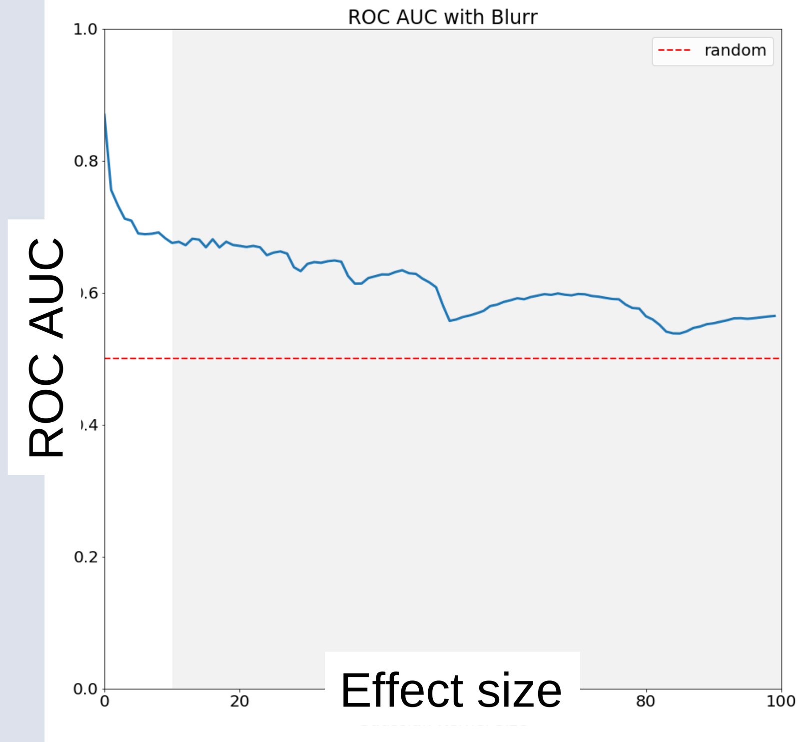


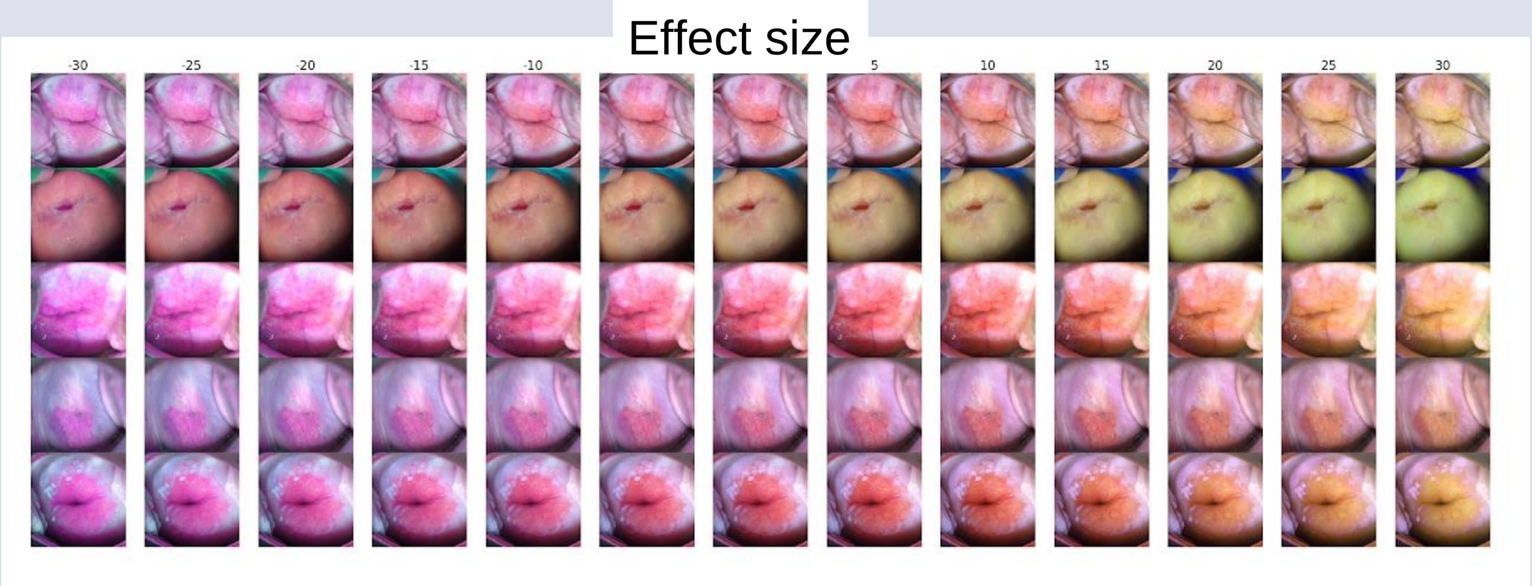


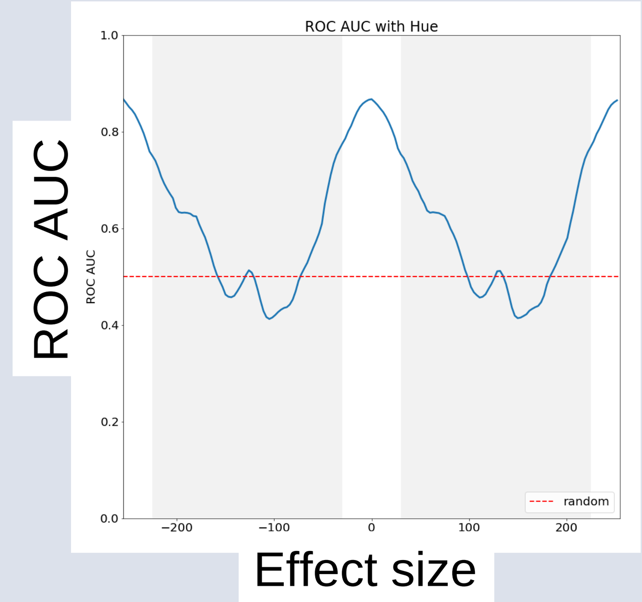
Results

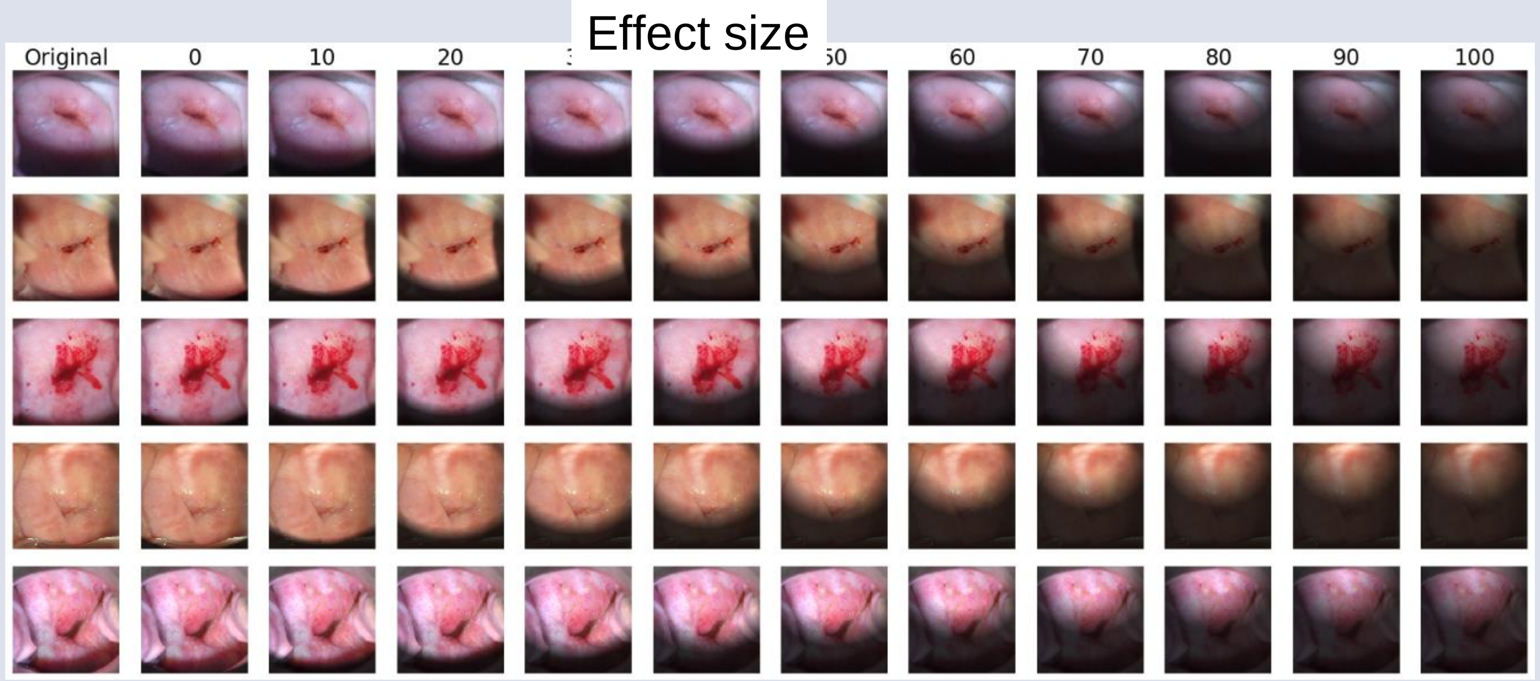
The 3 most relevant augmentation analyses are shown below. On the left are examples of an effect of different sizes on a handful of images. On the right is the corresponding drop in the area under the (ROC – Receiver Operating Characteristic) curve (AUC). Note the effect chart on left corresponds to white area in graphs. Gray regions correspond to images far outside the expected range, with little information.

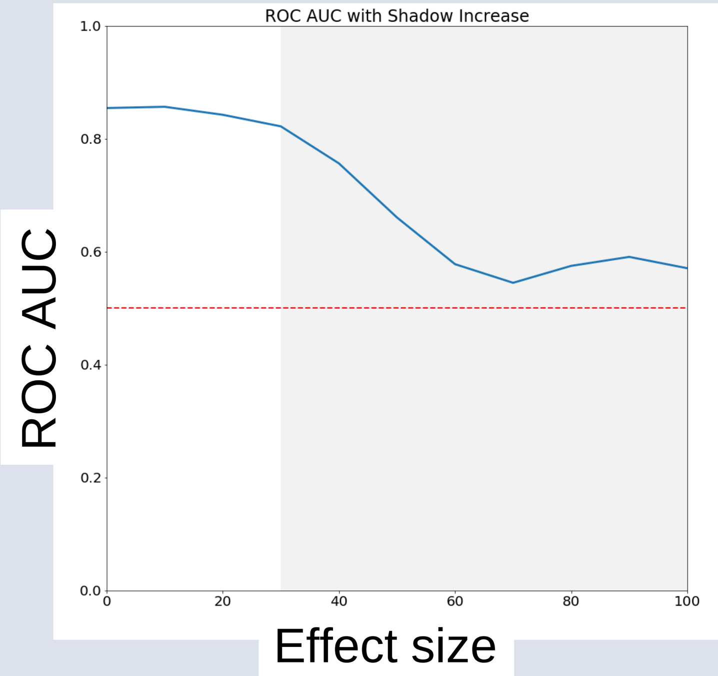












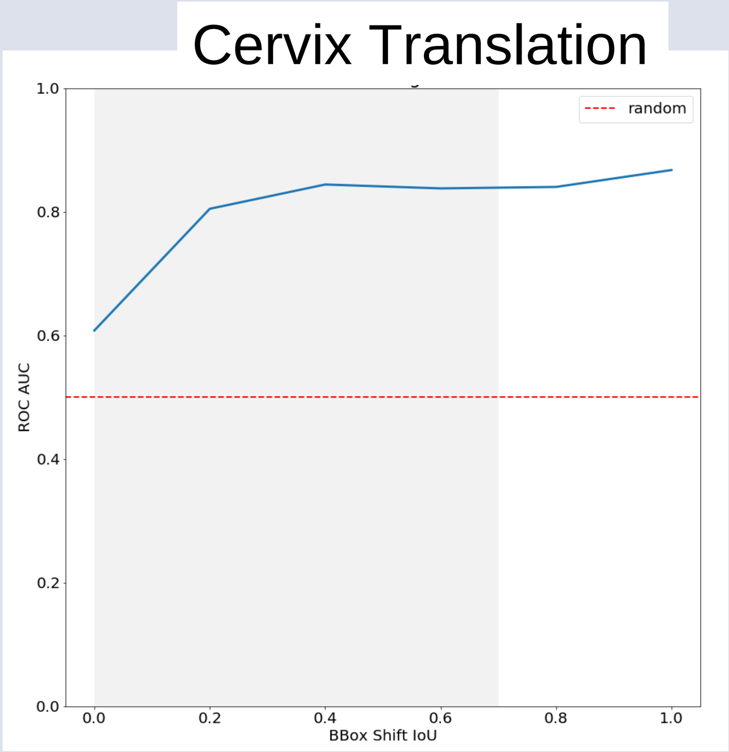
Device Impact on Machine Learning Classifier Accuracy in Detecting Cervical Dysplasia

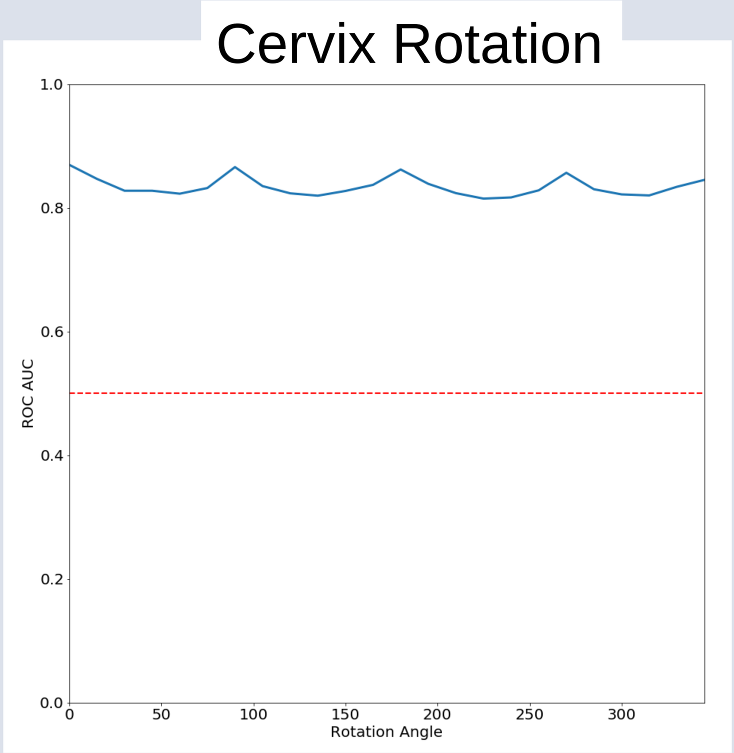
KC Fernandes[1], T Freitas[1], Y Zall[2], R Nissim[2], D Levitz[2]

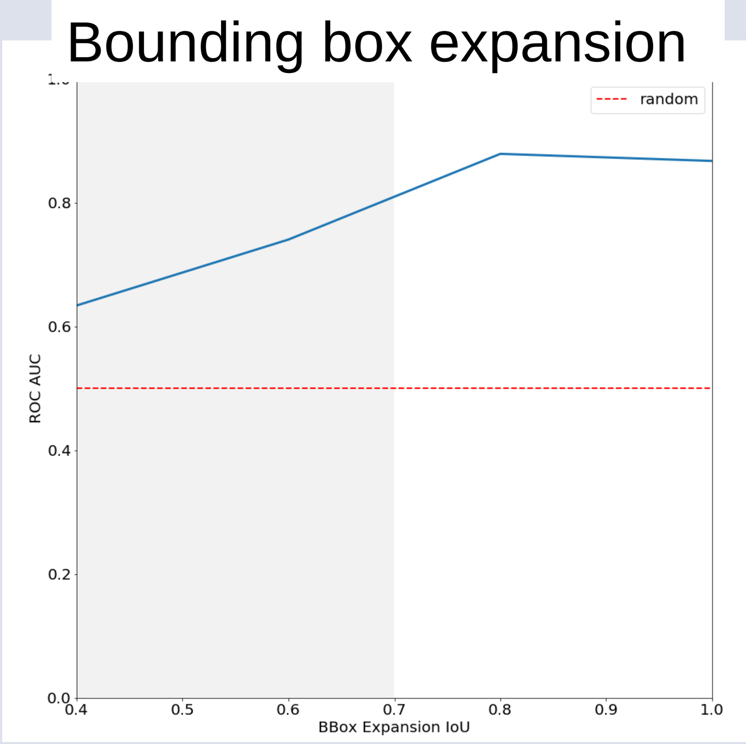
1 NILG.ai, Porto, Portugal; 2 MobileODT Ltd., Tel Aviv, Israel

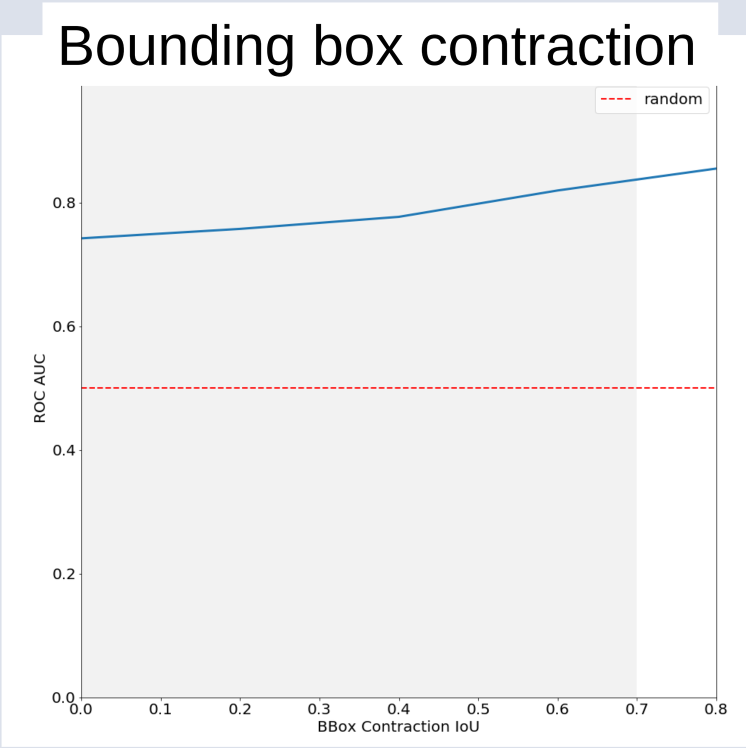
Additional augmentation analyses

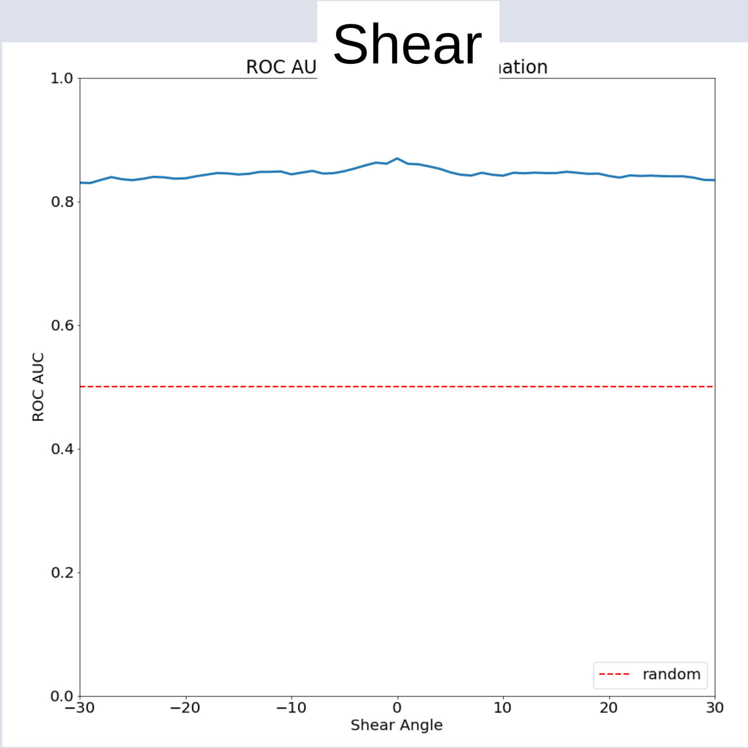
Augmentation analyses showing affine transformations did not influence the AVE prediction scores

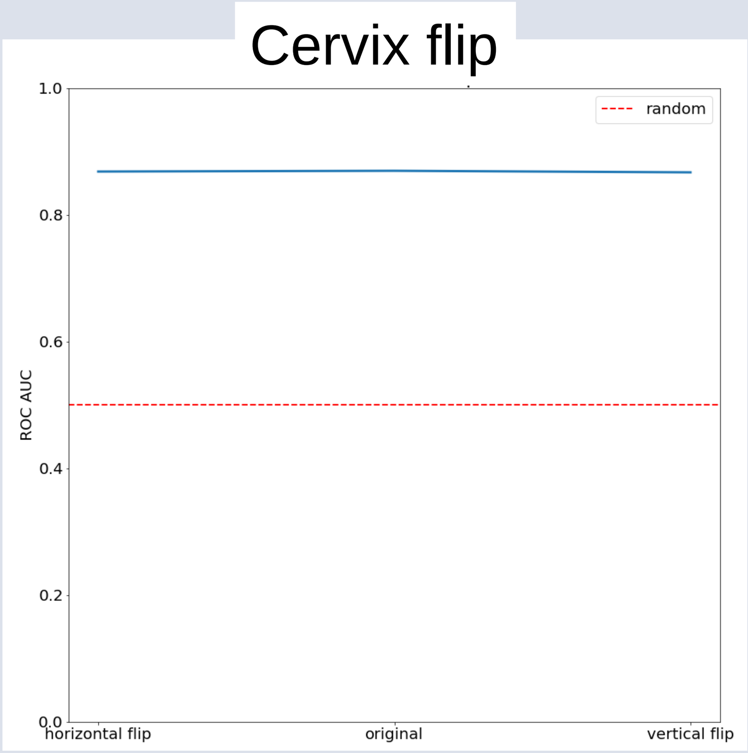






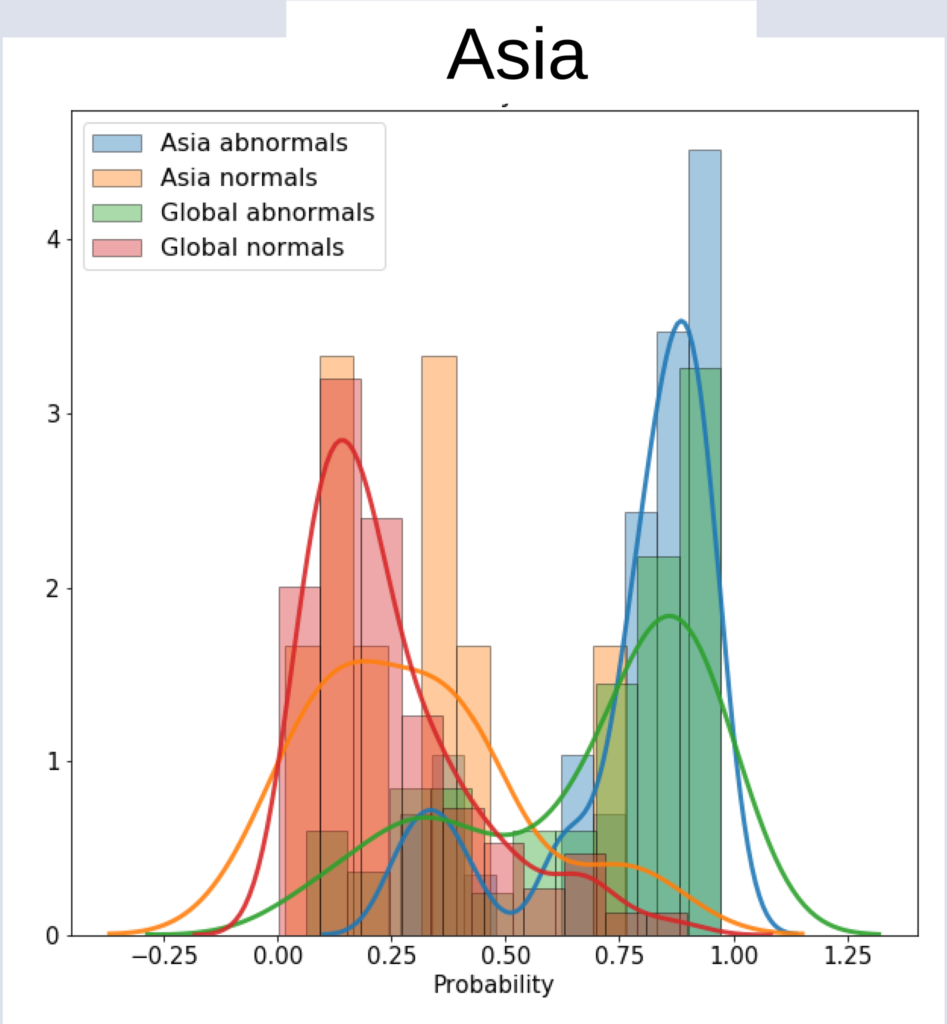


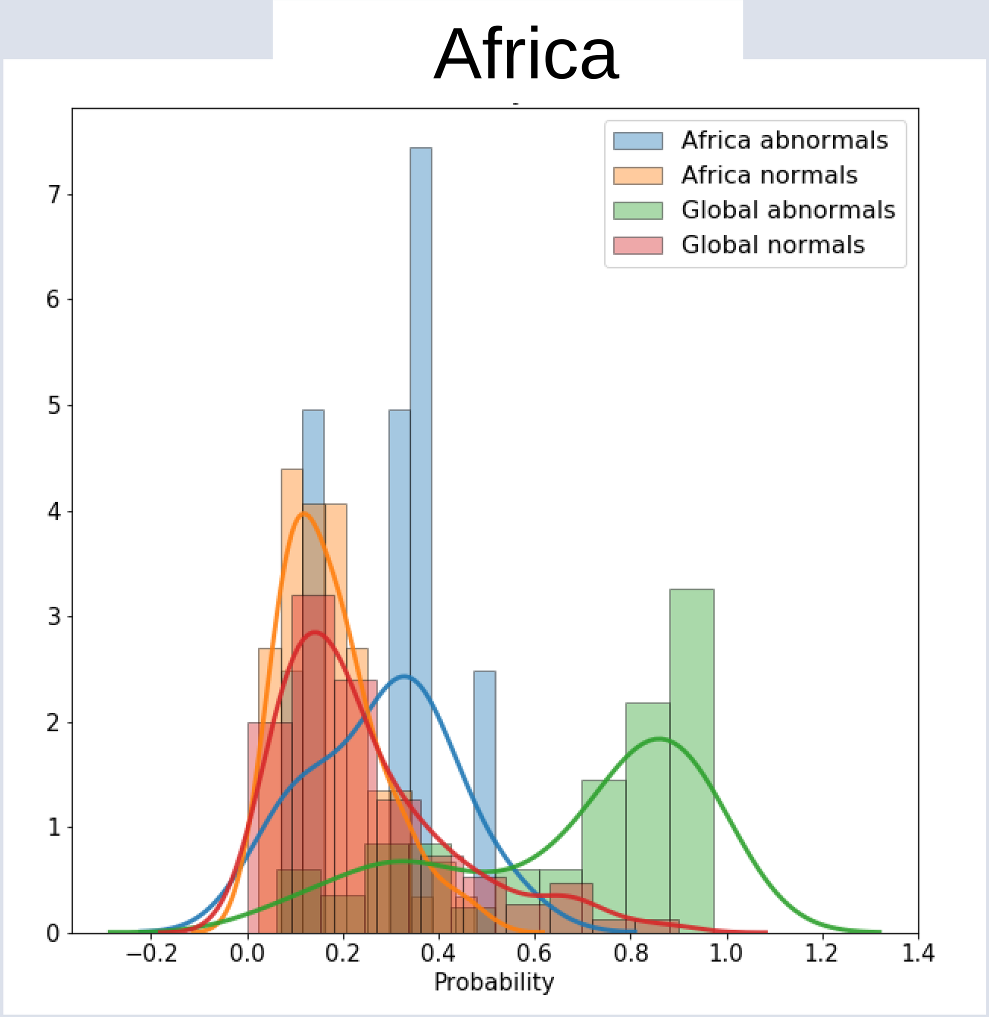


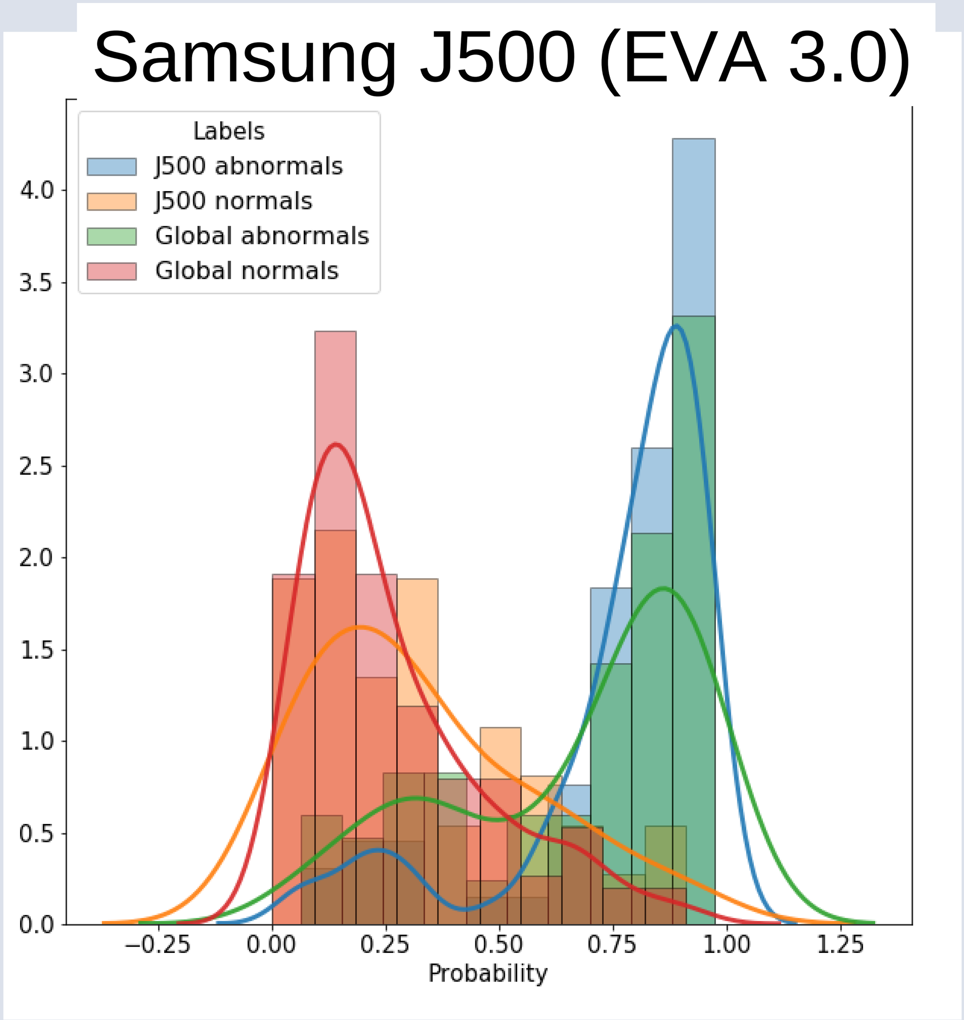


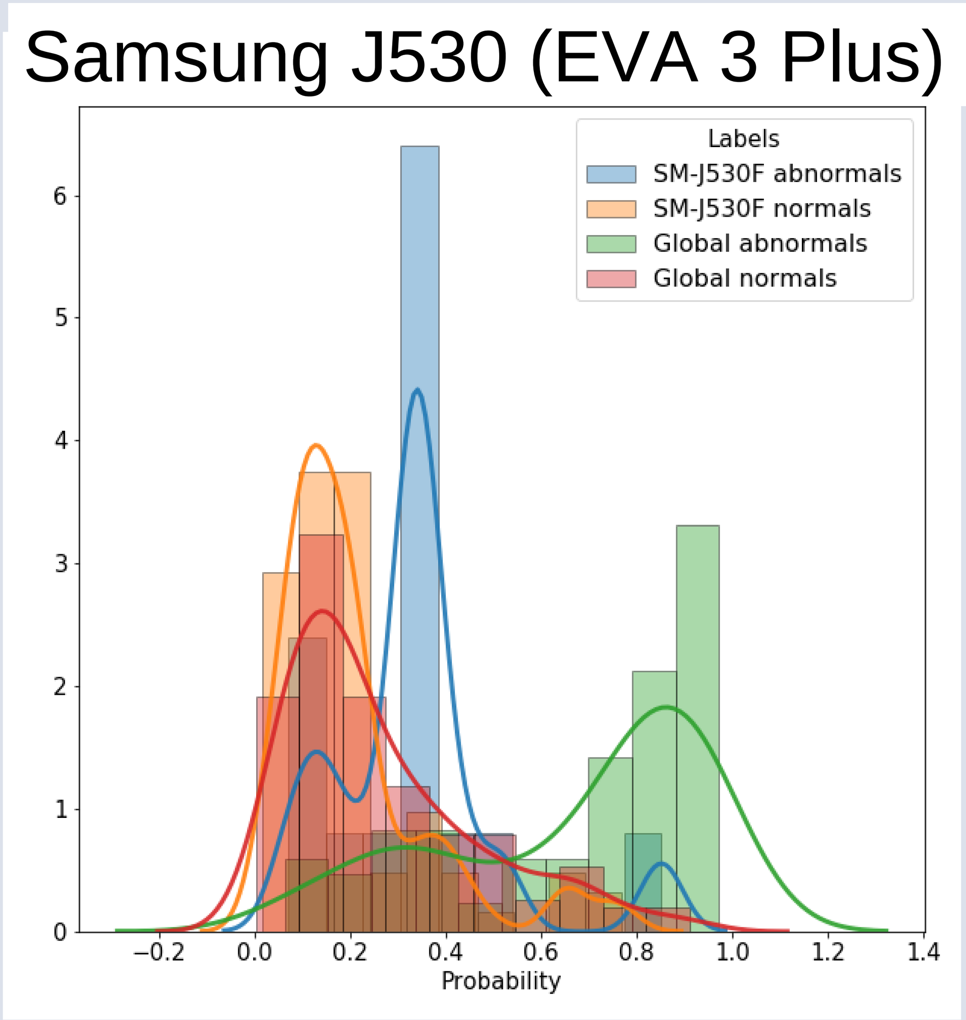
Statistical distribution analyses

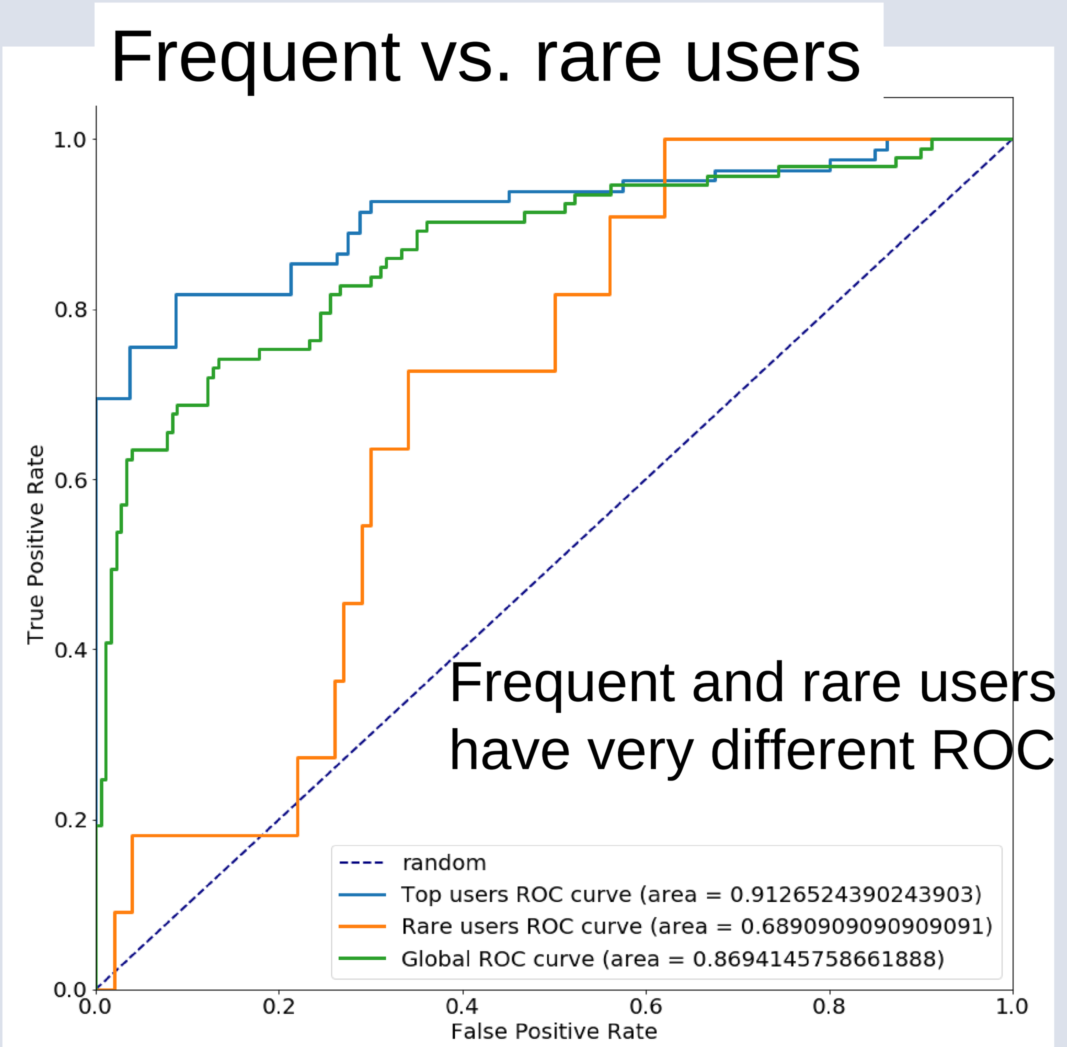
A comparison of the probability density functions (PDFs) of AVE prediction scores for different geographies and phone models. These results suggest that the classifier distinguishes between normal, and 2 types of abnormals – those from Africa (Samsung J530) and those from Asia (Samsung J500). Moreover, this suggests a correlation between phone model and geography











Frequent and rare users also have very different ROC curves.

CONCLUSIONS

The AVE classifier tested is very sensitive to blur, moderately sensitive to background lighting and shadows, and not sensitive to translations, rotations, scaling, shear, and flips. These results were expected.

The AVE classifier is also sensitive to phone model and geography. However, the 2 parameters cannot be isolated in the current data set. Additionally, the AVE classifier is sensitive to the frequency of usage, with better performance on common users, as opposed to rate users.

These tests and results should be considered when training and testing new AVE classifiers.

Funding

This study was funded by MobileODT.