Added value of feature uncertainty in the radiomic analysis of Contrast-Enhanced Digital Mammography images

Ricardo Montoya-del-Angel[a], Jorge Patricio Castillo-Lopez[b*], Vyanka Eunice Sanchez-Goytia[b], Liliana Moreno-Astudillo[b], Yolanda Villaseñor-Navarro[b], and Maria Ester Brandan[c]

[a]Facultad de Ciencias, UNAM, Mexico City, Mexico
[b]Instituto Nacional de Cancerología, Mexico City, Mexico.
[c]Instituto de Física, UNAM, Mexico City, Mexico

Abstract

Purpose: Radiomic analysis has shown the potential to improve clinical decision support systems. One of the challenges facing the clinical implementation of radiomic is to be reproducible. Our purpose is to show that assessing radiomic feature uncertainty can improve the robustness and performance of the radiomic analysis of contrast-enhanced digital mammography images. **Approach**: The prediction goal studied was the immunohistochemical status (IHC) of breast cancer in 33 patients. We consider a radiomic feature (RF) to be sensible to the presence of IHC antigens if it was selected in any of the prediction models. We assessed three sources of uncertainty: the mammography unit, misalignment between the subtracted images, and region-of-interest (ROI) delineation. The first two sources were evaluated in phantom studies, while the latter was quantified using multiple manual segmentations. **Results**: The uncertainty source with the largest influence on the RF values was the ROI segmentation, followed by image misalignment. Sensible radiomic features had an average uncertainty of 16%, ranging between 0.1% and 40%. Including uncertainties in the feature selection and training step significantly improved the performance of the prediction models. Specifically, for estrogen receptors, progesterone receptors, and Ki67, the Matthews Correlation Coefficient increased from 25%, 21%, and 26% to 43%, 42%, and 42%, respectively. Those same models achieved areas under the ROC curves of 77%, 79%, and 80%, respectively. **Conclusions**: Uncertainty used as a metric of RF robustness could improve radiomic workflow and model performance.

Keywords: radiomics, uncertainty, contrast-enhanced digital mammography.

***,** E-mail: jorge.castillo.mex@posgrado.unam.mx

1 Introduction

In 2020, breast cancer was the form of cancer with the highest estimated number of new cases for women, almost 25% of all new cases. In addition, 85% of countries reported by WHO had breast cancer as top cancer in terms of the age-standardized incidence rates in women[1]. Breast imaging plays an important role in the detection and diagnosis of breast cancer, and it is a field in continuous evolution. Contrast-enhanced digital mammography (CEDM) emerges as an imaging modality that produces images that express exclusively vascular contrast-uptake, increasing conspicuity and suppressing in the process most of the breast parenchyma[2].

Radiomics has arisen as a study field for the extraction of quantitative information from medical images and the mining of these data to create correlations with the biomedical state of the patients[3]. For this purpose, quantitative information descriptors, known as radiomic features (RF), could be used to develop aid tools to support clinical decision-making. Specifically for mammography, computer-aided detection and diagnosis (CAD and CADx) have become the two principal artificial intelligence techniques applied for this goal[4].

One of the biggest challenges faced in quantitative imaging is the lack of standardization and harmonization of the image acquisition, processing, and analysis[3]. These are considered non-biological factors that affect the image and the quantitative information extracted from it. Different approaches have been taken to consider RF variability, one of them being the assessment of uncertainty over the RF and the subsequent CADx model. Mendel et al.[5] conducted a study to quantify feature robustness across different mammography unit manufacturers, and Castillo-Lopez et al.[6] investigated the potential of uncertainty analysis to improve the evaluation of texture and the appraisal of the prediction model from CEDM single-energy images. Although some works have discussed the sources and effects of uncertainty, possible solutions to consider their effect have not yet been fully explored.

To this end, the purpose of this investigation is to show that the assessment of RF uncertainty can improve the robustness and performance of CADx prediction models. A radiomic analysis was designed to develop a prediction model correlating RF, from CEDM images in its single-energy (SET) modality, with the immunohistochemical (IHC) status of a breast lesion. Once uncertainty is quantified and its influence in the model prediction performance is evaluated, a restructuration of the model workflow is performed, which considers RF variation in the model training to preserve model robustness.

2 Materials and methods

2.1 Uncertainties in mammography images

A mammography image can be interpreted as a spatial distribution of X-ray attenuation measurements; these data present spatial correlations and uncertainty. Under this scope, uncertainty is inherited to the extracted RF from the image. RF uncertainty sources can be estimated considering the steps of the radiomic analysis workflow. In this investigation, due to the CEDM image acquisition and processing, we have considered the mammography unit, image misalignment, and region-of-interest (ROI) delineation as the three main uncertainty sources. In the CEDM context, image misalignment refers to misalignment of pre- and postcontrast medium images due to patient movement during image acquisition.

To assess the first two uncertainty sources, images of an experimental arrangement using a heterogeneous mixture of materials to simulate glandular and adipose tissue of the breast (using the *CIRS model 020 phantom)* were acquired. To simulate pre- and postcontrast CEDM, resins containing 0.0 mg/cm^2 and 3.0 mg/cm^2 of iodine, respectively, were set and exchanged over the surface of the phantom. Uncertainty due solely to the mammography unit was evaluated acquiring 10 pairs of pre- and postcontrast images, where each image pair was obtained keeping the phantom and resin positions unaltered. Uncertainty due to misalignment was isolated using 10 groups of mammography images. Firstly, 5 “precontrast” images were taken with no alteration between

them, then the phantom was slightly displaced and, secondly, 5 "postcontrast" images were acquired. The amount of displacement matched the one observed clinically. This process was repeated for different displacements thus resulting in different RF bias for each group, where bias was defined as the difference between the average value of RF with and without misalignment.

To avoid differences in scales among RF, uncertainty in RF was reported using the coefficient of variation (C_v), defined as

$$C_v = \frac{\sigma}{\mu}, \quad (1)$$

where σ is the sample standard deviation and μ the mean value. For the RF with a normalized scale, σ was used instead. For simplicity, in further notation C_v will be used for both cases and clarification will be made when needed.

Uncertainty due to ROI delineation was assessed using manual lesion segmentations by three radiologists. The image visualization tool *3D Slicer*[7–9] was used for ROI segmentation. Two delineation criteria were used to segment the lesion around its border, and to segment a focal ROI where there was high confidence of the presence of the lesion. Each radiologist repeated the process eight weeks after the first segmentation to include intra-observer variability. For the C_v value, an average of the 6 border segmentations, and their standard deviation, was used to include inter- and intra-radiologist variability.

2.2 Data acquisition

Forty-five SET studies from a previous research protocol were retrospectively reviewed.[10] The research protocol objective was to assess multicentric breast cancer using CEDM. Biopsy of the suspicious lesions were obtained for all patients and the IHC analysis included the status of estrogen receptors (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki67 biomarker for proliferation rate. Ki67 percentage score higher than 20% was assigned to positive status, which indicates a high proliferation rate. Eight studies were excluded due to presence of artifacts in the images, considered to affect the study result, and other four because of their benign lesion diagnosis. The distribution of positive cases of the ER, PR, HER2 and Ki67 analysis for the remaining 33 patients was as follow: 18%, 28%, 15% and 48% (Ki67>0.2), respectively.

The SET study clinical protocol included a basal mammography, a contrast agent injection, and four consecutive mammograms after injection, one minute apart. An affine transformation was used to align the pre- and postcontrast images. In this work, we used the 1-minute SET images because at this moment the misalignment artifacts would be minimal in comparison to the other images. The same mammography unit, a Fujifilm Amulet Innovality, was used for this clinical protocol and for the phantom images described in section 2.1. The settings for the acquisitions were W anode, Rh filter, and 45 kV + 5 mm external Al filter. Images were saved as DICOM "for processing" files, $100\ \mu m$ pixel size. A Fujifilm intrinsic software-based image preprocessing, which included a logarithmic response function and a flat-panel correction, was part of the "for processing" files.

2.3 *Impact of the radiomic features uncertainty on the prediction model*

Radiomic features were extracted using Pyradiomics, an open-source package in Python[11]. We considered 107 RF classified in three groups: shape-based, first-order statistics, and second-order

statistics. The extraction parameters were 64 grey level requantization, cooccurrence distance of three pixels, and a non-symmetrical definition of the Gray Level Coocurrence Matrix (GLCM). These parameters had been optimized for SET images in a previous work[12]. All RF were extracted on manually segmented regions of interest, as explained in section 2.1.

In general, building a prediction model consists of four stages: pre-processing, selection of the features with the greatest predictive power, training of the prediction model, and performance evaluation. In this work, we considered three different workflows to build the models to predict the IHC status (figure 1). The workflows were labeled as "traditional", "robust RF", and "robust training", considering how they included RF uncertainties. The traditional workflow did not consider RF uncertainties. The robust RF workflow prioritized the selection of features with lower uncertainty. The robust training workflow included the RF uncertainties in two stages: feature selection and model training. The models obtained from each workflow were compared in terms of the Matthews Correlation Coefficient (MCC)[13]. This coefficient is especially useful for evaluating the performance of a binary classifier when the data are not evenly distributed between the two classes. This is the case of the distribution of three antigens of the IHC status for the 33 studies.

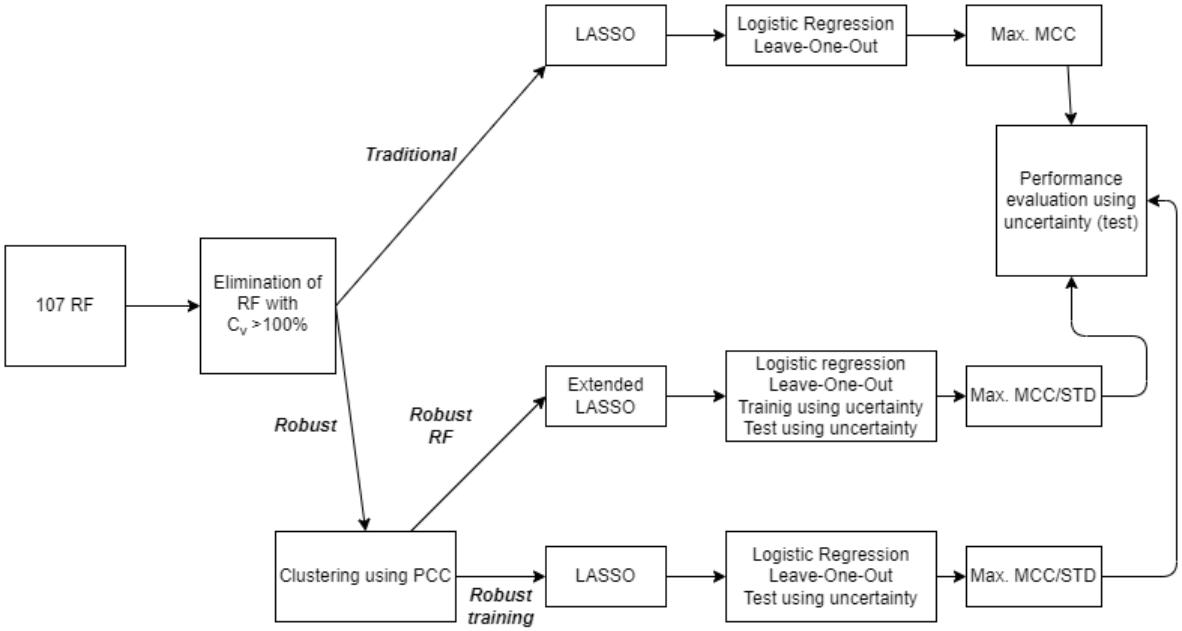

Figure 1. Flowchart of the radiomic analysis. The clustering process uses the Person Correlation Coefficient (PCC) as robust metric. A Least Absolute Shrinkage and Selection Operator (LASSO) analysis was used in the feature selection stage. Extended LASSO stands for the LASSO analysis using gaussian-generated RF values based on uncertainty. All workflows use Logistic regression for the model training.

All workflows shared the same preprocessing stage. First, RF whose C_v was greater than 100% were excluded. Next, the number of principal components needed to reproduce 99% of the RF variance was estimated using the principal component analysis (PCA) method. This value was used as a reference in the next stage to limit the number of RF that could be selected. For the two robust workflows, RF were clustered using Pearson's correlation coefficient (PCC) as distance metric, considering RF with PCC>0.95 to be correlated.

The selection of the RF with the highest predictive power was carried out using the LASSO method (least absolute shrinkage and selection operator). This technique was selected for its simplicity and because it has previously been used in the analysis of quantitative indicators of mammography with relatively small samples[14,15]. The parameter λ, which influences the number of indicators selected by the LASSO method, was adjusted considering the result of the PCA method. In this step, the robust workflows used the RF with the lowest relative uncertainty within each cluster, while the traditional workflow used all RF.

For each subset of the RF selected in the previous stage, a logistic regression was performed and the model with the highest MCC or $\frac{MCC}{\sigma_{MCC}}$ was identified. The MCC estimation was carried out by the leave-one-out method (LOO). In the first two workflows, traditional and robust RF, this stage was completed with the data of the 33 patients. On the other hand, in the robust training workflow the sample was increased considering the uncertainties of the RF. The procedure was as follows: for each contrasted mammography study, represented as an "n-dimensional" vector of RF (with n being the number of RF), another 1000 vectors were generated using a random kernel, with a n-dimensional Gaussian distribution (known as Gaussian disturbance), whose standard deviations were equal to the uncertainties measured for each RF. Through this procedure, the uncertainty information was included in the construction of the prediction models.

Finally, all workflows shared the same performance evaluation. The performance metric was the MCC calculated using leave-one-out. To avoid overoptimistic results and to include uncertainty, each evaluation was repeated 1000 times introducing random perturbations, with a Gaussian distribution, centered on the sample and with a standard deviation equal to the uncertainty of each indicator.

3 Results and discussion

3.1 Uncertainty distribution

Figure 1 summarizes the uncertainty distribution of the 99 RF with C_v<100%; it shows the influence of uncertainty sources (labeled by colors) on each RF. We can appreciate that the source with the largest influence on the RF uncertainty was the ROI delineation. Because image subtraction is a characteristic procedure for CEDM images, it is important to emphasize that first-order features showed the highest dependence on misalignment. In general, it was the second most influential uncertainty source for most of RF. Finally, uncertainty due to the mammography unit had the lowest influence on the uncertainty distribution, with an average contribution to the total C_v below 5%.

We can see that sensible RF are present in the three groups of RF types, with a clear majority of them in the second-order RF group. We consider a RF to be sensible to the presence of IHC antigens if it was used in any of the prediction models proposed in this work. Some of these RF have already been suggested to be sensible to IHC antigens in other investigations, as that of Zhou et al.[16], where 5 of these RF were found sensible to the presence of HER2 in breast cancer lesions. Mao et al.[14] and Marino et al.[17] also included the use of similar RF for their radiomic analysis although sensible RF were not explicitly reported.

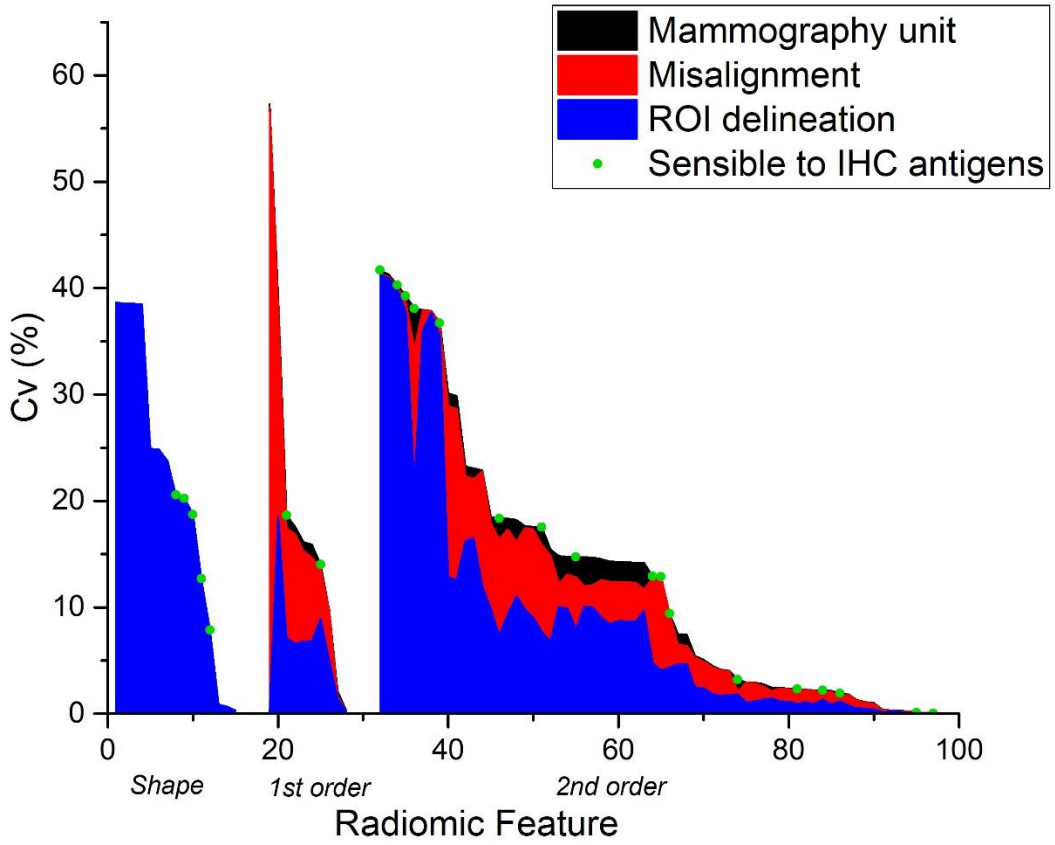

Figure 2. Uncertainty distribution of the three uncertainty sources for the 99 RF, in descendent order, with respect of total C_v. RF are separated by type in the three main groups. Green points represent the RF sensible to the presence of any of the four IHC antigens.

Table 1 shows the RF sensible to the presence of IHC antigens according to our classification model. HER2 is not included and will be discussed separately. Two RF showed sensibility to all prediction endpoints, namely *skewness* and *IMC2*. La Forgia et al.[18] found correlation between s*kewness* and PR receptor in contrast-enhanced spectral mammography. An interpretation of *skewness* is to consider it as a measure of symmetry in the contrast uptake, so the presence or absence of symmetry appears to be imperative for the characterization of breast lesions. On the other hand, IMC2, the Informational Measure of Correlation, was found correlated to the IHC status by Zhou et al.[16] and Castillo et al.[6], for Standard Mammography and CEDM, respectively. Both RF had a total C_v below 15% and were primarily influenced by image misalignment.

It is important to emphasize that low C_v values, which enable a RF to be considered robust and reproducible, do not imply sensibility towards IHC antigens. Evidence of this is the total uncertainty of all sensible RF found in this investigation, which ranges from about 0.1% to 40%. Previous works have stated that RF with high natural biologic variability will be the most informative.[19] In that sense, preference for RF with low C_v, as performed in this work, could separate interpatient variability due to the acquisition and measurement process from the actual biological variability.

Table 1. Uncertainty distribution of all RF sensible to the presence of immunohistochemical antigens in terms of C_v. 2nd-order RF are grouped according to their gray level matrix, whose description and definition can be found in the Pyradiomics documentation[11].

		Radiomic Feature	Mammography Unit	Misalignment	Region of interest (ROI)	Total	Immunohisto-chemical status
Shape	2	Maximum 2D Diameter Row	0.00%	0.00%	20.55%	20.55%	ER
	11	Minor Axis Length	0.00%	0.00%	18.73%	18.73%	ER
	15	2D Sphericity	0.00%	0.00%	7.85%	7.85%	ER
1st order	27	Range	6.71%	15.84%	7.19%	18.64%	PR
	30	Skewness	3.27%	10.35%	8.89%	14.04%	ER, PR, Ki67
2nd order GLCM	34	Cluster Prominence	16.64%	26.92%	21.21%	38.10%	Ki67
	38	Correlation	0.41%	12.20%	4.08%	12.87%	ER
	47	Imc2	1.49%	8.24%	4.32%	9.42%	ER, PR, Ki67
	52	MCC	2.13%	11.80%	4.78%	12.91%	PR
	55	Sum Entropy	1.10%	1.36%	1.34%	2.20%	Ki67
2nd order GLSZM	57	Gray Level Non-Uniformity	3.53%	4.33%	39.92%	40.31%	PR
	70	Zone Entropy	0.69%	1.98%	0.95%	2.30%	Ki67
2nd order GLRLM	79	Long Run Low Gray Level Emphasis	0.02%	0.03%	0.02%	0.04%	PR
	81	Run Entropy	1.04%	1.10%	1.16%	1.91%	PR
2nd order NGTDM	89	Busyness	6.70%	8.49%	37.75%	39.27%	ER
	91	Complexity	7.51%	13.84%	7.70%	17.53%	PR
	92	Contrast	2.41%	1.10%	1.81%	3.21%	ER
2nd order GLDM	95	Dependence Non-Uniformity	3.07%	9.01%	35.46%	36.71%	PR

First-order RF were the less sensible, which suggests that relevant information to predict a IHC status resides on the shape of the lesion and the spatial correlation of iodine absorption and not in iodine absorption per se. This contrasts with the finding of la Forgia et al.[20] who found correlation with IHC status using exclusively first-order features.

3.2 Prediction models comparison

The performance of each prediction model is represented in figure 3 using the median MCC and its interquartile range (IQR). MCC median values represent the prediction model performance, whereas its IQR is a measurement of robustness to variations of the data sample in the order of the RF uncertainties. We recall that the use of median and IQR is due to generation of new data for each testing sample, as explained for the LOO evaluation technique in section 2.4.

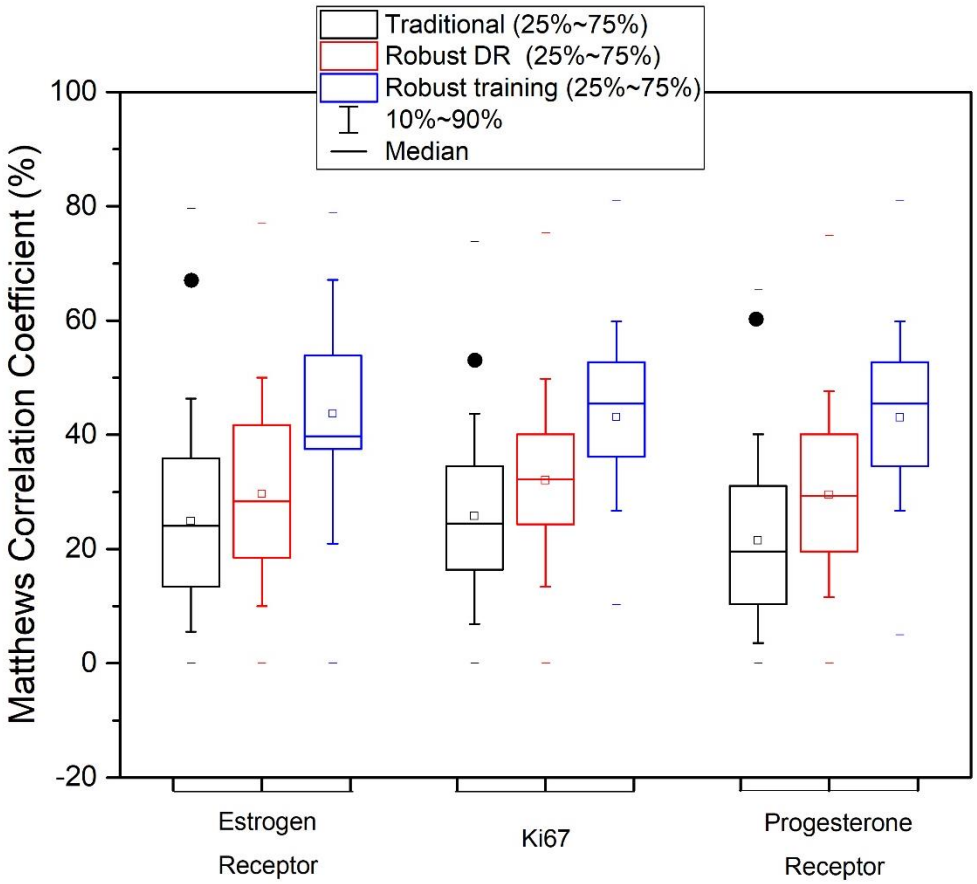

Figure 3. Models of the three workflows for each immunohistochemical antigen, except HER2. The black dots show the performance of the traditional model evaluated without introducing uncertainty perturbations.

The performance of traditional models dropped significantly when it was evaluated using synthetic data in the uncertainty range of the radiomic features. For instance, a model with an initial *moderate* agreement ($MCC = 67.7\%$) predicting ER status happened to have a *fair* agreement ($\mu_{MCC} = 24.9\%$) after tested using data augmentation considering feature uncertainties. The qualitative interpretation in terms of MCC values is taken from Cohen's[21] interpretation of Cohen's Kappa, a performance metric similar to MCC. The traditional prediction models for the rest of the antigens show an analogous behavior. This can be understood that if a clinical study were to be repeated with the same patients and exact acquisition conditions, the previously defined prediction model would reduce its performance for the newly acquired images for the same patients, failing to generalize its prediction and performance. A reason for this behavior could be that LASSO feature selection does not contemplate feature instability, so selected RF could be highly sensitive to uncertainty, compromising the model reproducibility. In any case, the results indicate that the initial overoptimistic-model performances were fictional due to overfitting or causality, and that its effects disappear when evaluated with the data augmentation testing technique. Determining uncertainty influence in prediction models is a problem that has previously been approached for different image modalities, including CEDM. For the latter, Castillo et al.[6] determined the influence of RF values, extracted from CEDM images, on prediction models for the same IHC antigens, using a Figure of merit (FOM) as performance metric.

Selecting robust RF from each cluster group allowed a slight improvement of MCC average values, as red boxes show in figure 3. Although all models improved their average MCC performance, the model agreement remained *fair* (below $\mu_{MCC} = 40\%$) for all antigens. This could be because models were not trained to recognize variation in the uncertainty range of the RF values, which could be expected if the clinical study were repeated. Clustering technique can be found in other investigations, such as the conducted by Robinson et al.[22], who grouped RF using the PCC as distance metric, as we did in this work. The way robustness is measured is an important term to be defined. Kalpathy-Cramer et al.[23] used the concordance correlation coefficient (CCC) as robustness metric, while we used uncertainty expressed in terms of the standard deviation.

Augmenting the training data using each feature uncertainty showed a considerable recovery of the model performance for 3 antigens, as showed in figure 3. The three models scaled to the *moderate* model agreement category (above $\mu_{MCC} = 40\%$). Specifically, for ER, PR and Ki67, MCC performance increased from 24.5%, 21.0%, 25.7% to 43.2%, 42.2% and 42.1%, respectively. This training proposal is important because such fluctuation of RF values is intrinsic for CEDM images due to its acquisition and processing protocol.

To ensure a statistically significant difference between means of the blue and red boxes in figure 3, a Mann-Whitney test was applied for each antigen. In all cases, a $p - value < 0.05$ was obtained, concluding that robust training models had statistically different means compered to robust DR selection models.

We include the ROC curve of the PR status prediction models for the three workflows, as shown in figure 4. The area under the curve (AUC) for the traditional, robust RF, and robust training model had confidence intervals of (61.5%, 63.6%, 65.5%), (67.5%, 69.3%, 70.9%) and (77.3%, 79.0%, 80.5%), respectively. The error bars were defined using a set of bootstrap replicas of the true positive rate values for each false positive rate value. For visualization purposes we show only 20 error bars. In average, AUC improved from 63.6% to 79% using RF uncertainties in the training of the model, a circa 15% AUC difference. The other two antigens prediction models, ER and Ki67, had a similar AUC improvement. These results are in agreement with the increase in the MCC presented in figure 3.

In figure 5 the ROC curves of the best models for ER, PR, and Ki67 are shown, where these best models are the robust training prediction models. The AUC confidence intervals were (75.1%, 77.4%, 79.7%) for ER, (77.3%, 79.0%, 80.5%) for PR, and (78.1%, 79.8%, 81.3%) for Ki67. These three models have AUC values around 80%, which are comparable to AUCs published in previous works. In their study to assess the relation between 1st- order RF and the IHC status, La Forgia et al.[20] obtained an AUC of 76% for the PR prediction model. Besides, Marino et al.[17] found an AUC between 74% and 78% for a PR or ER prediction model trained with shape, 1st-and 2nd- order RF.

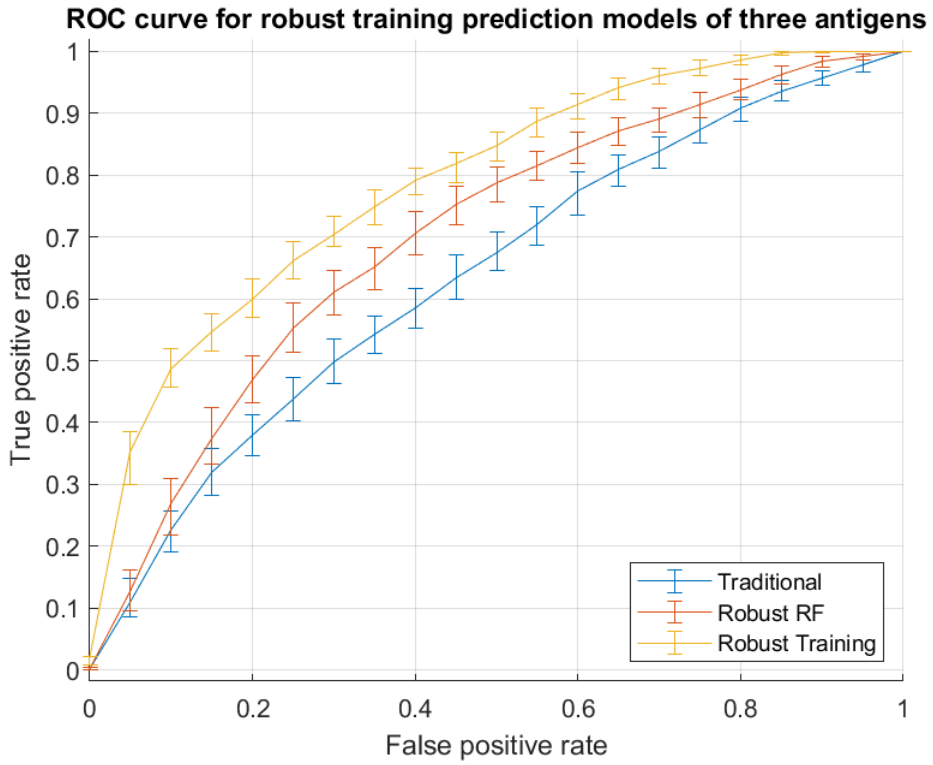

Figure 4. ROC curves of the progesterone receptor prediction model for the traditional, robust feature and robust training workflows. Average area under the curve of 63.6%, 69.9%, and 79%, respectively.

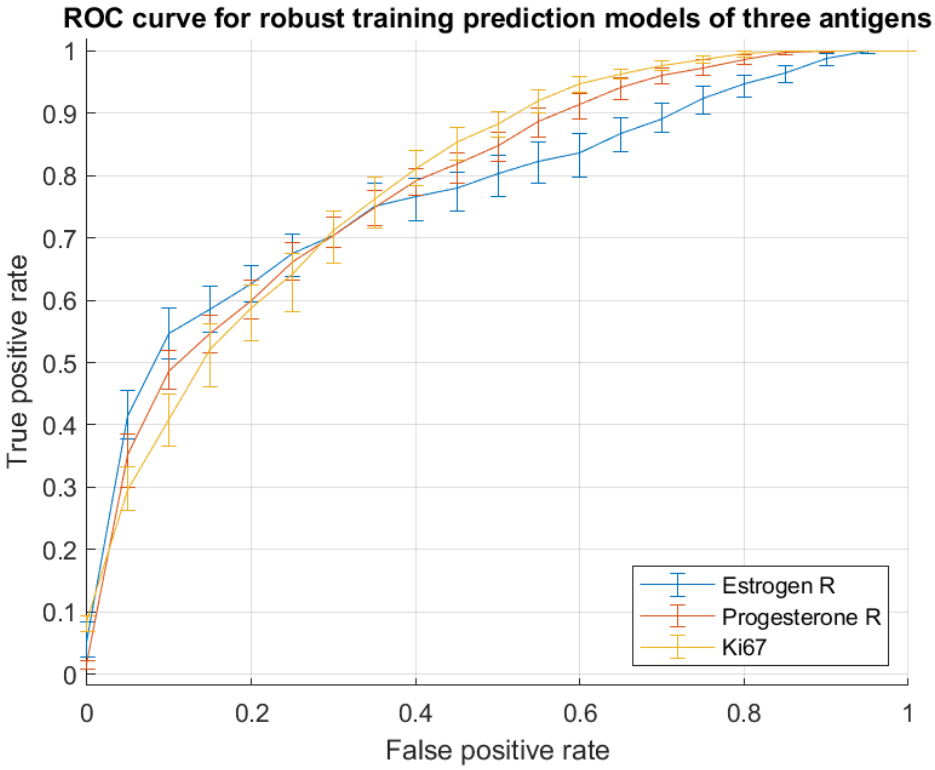

Figure 5. ROC curves of Ki67, progesterone and estrogen receptors. Average AUC of 79.8%, 79% and 77.4%, respectively.

The model obtained from the traditional workflow to predict HER2 status had lower MCC values than those from robust workflows (21.9% vs 31.9% and 32.2%). However, HER2 was the only antigen whose models did not improve with robust training (just 0.03% increase in the median MCC). We believe that the lack of improvement could be because HER2 do not have a solid relation with this RF sample. This does not imply that the presented RF have no relation with HER2 presence at all. As Zhou et al. showed, there exist a sensibility to the presence of HER2 for 5 out of 9 RF used in this work, although their worked included standard mammography and not CEDM. This last consideration may suggest that anatomical information could be better correlated that functional. The lack of sensibility found could be overcome by increasing the number of samples of the training set.

For further comparison with other works, a set of 4 tables containing the performances of all mentioned models in terms of the traditional performance metrics, i.e., F1 score, precision, and accuracy, can be found in the appendix.

4 Conclusion

The progress of Radiomics has brought up the need of standardization of radiomic analysis and the detailing of its workflows. This may include the consideration of uncertainty as an important part of the development and its application of this type of analysis, such as computer-aided tools for diagnosis and detection. To this end, we've proposed a technique for the evaluation of the prediction performance of models correlating RF and the presence of IHC antigens from CEDM images, in its SET modality, including a detailed workflow for the feature extraction, model training and testing. We've compared the importance of the inclusion of uncertainty on each of the three mentioned steps, defining different model approaches for each antigen prediction. We've found that the traditional models, which did not included uncertainty on their workflow, had a MCC reduction from around 30 and 40%, losing considerable prediction performance. An increase from 5 to 10% of MCC was reached after considering robust RF in the feature selection. A final MCC value above 40% and an AUC of about 80% were obtained for ER, PR and Ki67>0.2 prediction models after considering uncertainty variability of the RF values in the training set. Results showed that the uncertainty source with the highest influence on the RF values was the ROI segmentation, followed by image spatial registration. In contrast, uncertainty used as a measurement of RF robustness and included in the radiomic workflow can improve the RF selection and optimize the prediction model performance.

Disclosures

The authors have no financial disclosures.

Acknowledgments

We gratefully acknowledge partial financial support from Conacyt-Mexico Grant 1311307 "Imágenes radiológicas cuantitativas para la caracterización no invasiva del cáncer de mama",

Fundación Miguel Alemán A.C., and PAPIIT-UNAM grant IN-103219. We acknowledge Martha Vargas Gayón, MD, for her contribution to lesion segmentation.

References

[1] IARC., "New Global Cancer Data: GLOBOCAN 2020," 2020.

[2] Lobbes, M. and Jochelson, M., [Contrast-Enhanced Mammography, 1st ed.], Spring International Publishing (2019).

[3] Lambin, P., Leijenaar, R. T. H., Deist, T. M., Peerlings, J., De Jong, E. E. C., Van Timmeren, J., Sanduleanu, S., Larue, R. T. H. M., Even, A. J. G., Jochems, A., Van Wijk, Y., Woodruff, H., Van Soest, J., Lustberg, T., Roelofs, E., Van Elmpt, W., Dekker, A., Mottaghy, F. M., Wildberger, J. E., et al., "Radiomics: The bridge between medical imaging and personalized medicine," Nat. Rev. Clin. Oncol. **14**(12), 749–762 (2017).

[4] Katzen, J. and Dodelzon, K., "A review of computer aided detection in mammography," Clin. Imaging **52**(August), 305–309 (2018).

[5] Mendel, K. R., Li, H., Lan, L., Cahill, C. M., Rael, V., Abe, H. and Giger, M. L., "Quantitative texture analysis: robustness of radiomics across two digital mammography manufacturers' systems," J. Med. Imaging **5**(01), 1 (2017).

[6] Castillo Lopez, J. P., Montoya-del-Angel, R., Sanchez Goytia, V., Moreno Astudillo, L., Villaseñor Navarro, Y. and Brandan, M.-E., "Uncertainties associated to the extraction of texture features in single-energy contrast-enhanced mammography," 45 (2020).

[7] Kikinis, R., Pieper, S. D. and Vosburgh, K. G., "3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support," Intraoperative Imaging Image-Guided Ther., 277–289 (2014).

[8] Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J. V., Pieper, S. and Kikinis, R., "3D Slicer as an image computing platform for the Quantitative Imaging Network," Magn. Reson. Imaging **30**(9), 1323–1341 (2012).

[9] Pieper, S., Fillion-Robin, J. C. and Kikinis, R., “3D Slicer,” 2015.

[10] Sánchez Goytia, V. E., “Detección de lesiones secundarias de mama usando mastografía con medio de contraste” (2019).

[11] Van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillion-Robin, J. C., Pieper, S. and Aerts, H. J. W. L., “Computational radiomics system to decode the radiographic phenotype,” Cancer Res. **77**(21), e104–e107 (2017).

[12] Mateos, M. J., Gastelum, A., Márquez, J. and Brandan, M. E., “Texture analysis of contrast-enhanced digital mammography (CEDM) images,” Breast Imaging. 13th Int. Work. IWDM 2016 Malmö, Sweden, June 2016, Proceedings., 585–592 (2016).

[13] Chicco, D. and Jurman, G., “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” BMC Genomics **21**(1), 1–13 (2020).

[14] Mao, N., Yin, P., Li, Q., Wang, Q., Liu, M., Dong, J., Che, K., Wang, Z., Duan, S., Zhang, X., Xie, H. and Hong, N., “Radiomics nomogram of contrast-enhanced spectral mammography for prediction of axillary lymph node metastasis in breast cancer: a multicenter study,” EBioMedicine (2019).

[15] Fusco, R., Piccirillo, A., Sansone, M., Granata, V., Rubulotta, M. R., Petrosino, T., Barretta, M. L., Vallone, P., Di Giacomo, R., Esposito, E., Di Bonito, M. and Petrillo, A., “Radiomics and Artificial Intelligence Analysis with Textural Metrics Extracted by Contrast-Enhanced Mammography in the Breast Lesions Classification,” Diagnostics **11**(5), 815 (2021).

[16] Zhou, J., Tan, H., Bai, Y., Li, J., Lu, Q., Chen, R., Zhang, M., Feng, Q. and Wang, M., “Evaluating the HER-2 status of breast cancer using mammography radiomics features,” Eur. J. Radiol. **121**(January), 108718 (2019).

[17] Marino, M. A., Pinker, K., Leithner, D., Sung, J., Avendano, D., Morris, E. A. and Jochelson, M., “Contrast-Enhanced Mammography and Radiomics Analysis for Noninvasive Breast Cancer Characterization: Initial Results,” Mol. Imaging Biol. **22**(3), 780–787 (2020).

[18] Losurdo, L., Fanizzi, A., Basile, T. M. A., Bellotti, R., Bottigli, U., Dentamaro, R., Didonna, V.,

Lorusso, V., Massafra, R., Tamborra, P., Tagliafico, A., Tangaro, S. and La Forgia, D., "Radiomics analysis on contrast-enhanced spectral mammography images for breast cancer diagnosis: A pilot study," Entropy **21**(11) (2019).

[19] Gillies, R. J., Kinahan, P. E. and Hricak, H., "Radiomics: Images are more than pictures, they are data," Radiology **278**(2), 563–577 (2016).

[20] La Forgia, D., Fanizzi, A., Campobasso, F., Bellotti, R., Didonna, V., Lorusso, V., Moschetta, M., Massafra, R., Tamborra, P., Tangaro, S., Telegrafo, M., Pastena, M. I. and Zito, A., "Radiomic analysis in contrast-enhanced spectral mammography for predicting breast cancer histological outcome," Diagnostics **10**(9), 708–719 (2020).

[21] McHugh, M. L., "Lessons in biostatistics interrater reliability : the kappa statistic," Biochem. Medica **22**(3), 276–282 (2012).

[22] Robinson, K., Li, H., Lan, L., Schacht, D. and Giger, M., "Radiomics robustness assessment and classification evaluation: A two-stage method demonstrated on multivendor FFDM," Med. Phys. **46**(5), 2145–2156 (2019).

[23] Kalpathy-cramer, J., Mamomov, A., Zhao, B., Lu, L., Cherezov, D., Napel, S., Echegaray, S., Rubin, D., Mcnitt-gray, M., Lo, P., Sieren, J. C., Uthoff, J., Dilger, S. K. N., Driscoll, B., Yeung, I., Hadjiiski, L., Cha, K., Balagurunathan, Y., Gillies, R., et al., "Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features," Tomography **2**(4), 430–437 (2016).