AbstractID: 10510 Title: Exploring Non-Linear Feature Space Dimension Reduction and Data Representation in Breast CADx

Purpose: Preliminary study to explore potential of recently developed non-linear dimension reduction and data representation techniques as applied to breast lesion CADx generated feature spaces. **Method and Materials:** Two new methods were explored: Laplacian eignenmaps of (Belkin and Niyogi) and t-distributed stochastic neighbor embedding (t-SNE) (van der Maaten and Hinton). The properties of these methods were evaluated in the context of malignancy classification performance as well as visual inspection of the sparseness for two and three dimensional mapping representations. The robustness of the proposed techniques were tested against four separate imaging modality feature databases, including 2956 ultrasound(US) , 735 full-field digital mammography (FFDM), 850 screen-film mammography (SFM), and 356 DCE-MRI biopsy proven breast mass lesions. Using the reduced dimension mapped feature output as input, as opposed to feature selection, two classifiers, linear and non-linear, were tested: Markov Chain Monte Carlo based Bayesian artificial neural network (MCMC-BANN) and linear discriminate analysis (LDA). To evaluate performance, the AUC was estimated for each classifier using ROC analysis and 0.632+ bootstrap validation on over 500 samples. In addition, performance was compared to a non-linear Automatic Relevance Determination (ARD), linear step-wise feature selection method, as well as a linear reduction Principle Component Analysis based method. **Results:** The new methods were found to match or exceed performance of current state-of-the-art breast lesion CADx algorithms for feature selection and classification across all modalities. Additionally, the new techniques possess the added benefit of naturally delivering sparse lower dimensional representations for visual interpretation, thus, revealing intricate data structure of the feature space. **Conclusion:** These methods maintain predictive power while also preserving both local and global structural information present in the original high dimensional feature space for breast lesion feature data used in CADx. **Conflict of Interest:** M. Giger is a stockholder and receives royalties from Hologic.