Breast Cancer Diagnosis in Digital Breast Tomosynthesis: Effects of Training Sample Size on Multi-Stage Transfer Learning using Deep Neural Nets

Supplementary Material

Ravi K. Samala, Heang-Ping Chan, Lubomir Hadjiiski,
Mark A. Helvie, Caleb D. Richer, Kenny H. Cha

VI. Statistical analysis

For the three sets of schemes shown in figures 5(d) and 6(d), two-tailed paired *t*-tests were performed to assess the statistical significance of the difference between pairs of transfer learning schemes at the corresponding training sample size.

Table V. *p*-values from two-tailed paired *t*-test between schemes shown in fig. 5(d) and fig. 6(d) at different training sample sizes. *p*-values less than 0.05 are considered statistically significant and highlighted in boldface and the rest are shown in blue color.

Sample size	Fig. 5(d)			Fig. 6(d)		
Percent of training data	**A, B**	**A, C**	**B, C**	**D, B**	**D, C**	**B, C**
1%	**< 0.05**	**< 0.05**	**< 0.05**	**< 0.05**	**< 0.05**	0.09
5%	**< 0.05**	**< 0.05**	**< 0.05**	**< 0.05**	**< 0.05**	0.10
10%	**< 0.05**	0.24	**< 0.05**	**< 0.05**	**< 0.05**	0.10
20%	**< 0.05**	0.22	**< 0.05**	**< 0.05**	**< 0.05**	0.06
40%	**< 0.05**	0.06	0.21	**< 0.05**	**< 0.05**	0.11
60%	**< 0.05**	0.06	0.80	**< 0.05**	**< 0.05**	0.35
80%	**< 0.05**	0.06	0.34	**< 0.05**	**< 0.05**	0.45
100%	**< 0.05**	**< 0.05**	0.06	**< 0.05**	**< 0.05**	**< 0.05**

This work is supported by National Institutes of Health award numbers R01 CA151443 and R01 CA214981
RK Samala, HP Chan, L Hadjiiski, MA Helvie, CD Richter, KH Cha are with the Department of Radiology, University of Michigan, Ann Arbor, MI. (e-mail: {rsamala,chanhp,lhadjisk,mahelvie,calebric,heekon} @umich.edu)

VII. Single- vs multi-stage transfer learning at matched number of samples

In *Scheme B*, 100% training data would include 2,242 mammography views at Stage 1 and 230 DBT views for Stage 2 (multi-stage). In *Scheme D*, 100% training data would include 230 DBT views for Stage 1 (single-stage). *Scheme B* would use a maximum (at 100%) of 2,242 additional mammography views for DCNN training. Note that the total number of unique ROIs from 230 DBT views was 1,140 since five slices per lesion were extracted. The two schemes may be compared at total number of views or total number of unique ROIs. Table VI below shows the distribution at different percentages of the training data for the two stages of transfer learning. Specifically, two conditions were highlighted that would match the sample size closely for the single-stage and two-stage training, where (i) by views: single stage at 100% of DBT views (230) can be compared with two stages at 5% of stage 1 mammography and 50% of stage 2 DBT, and (ii) by unique ROIs: single stage at 100% of DBT ROIs (1,140) compared with 25% of stage 1 mammography and 50% of stage 2 DBT. A range of scenarios were also compared to illustrate trends of varying the sample sizes as shown in Fig. 10 below.

The mean values at single-stage DBT (100%), single-stage DBT (50%) and multi-stage transfer learning (DBT 100% and mammography 100%) are 0.822, 0.724 and 0.886, respectively. The mean value at multi-stage learning starts from 0.818 at (MAM 5%, DBT 50%), increases to 0.837 at (MAM 25%, DBT 50%) and ends with 0.843 at (MAM 30%, DBT 50%). With equal number of views between the two approaches, the mean AUCs are comparable (0.822 vs 0.818). With equal number of unique ROIs, the mean AUC of multi-stage approach (0.837) is slightly greater than that of single-stage approach (0.822). More importantly, these and the results in Fig. 5 and Fig. 6 indicate that, when the available data from the target domain (e.g., DBT 50% in Fig. 10) are limited, an additional stage of pre-training using data from a similar auxiliary domain (e.g., mammography) consistently provides an advantage (pink curve) over transfer learning directly from ImageNet to the target domain (red curve) because of the similarity of low-level features between the two imaging modalities.

Table VI. Single- and multi-stage transfer learning approaches at different sample sizes of mammography (MAM) and DBT training data.

	Stage 1				Stage 2				Total	
	Type	**#views**	**#ROIs**	**Percentage**	**Type**	**#views**	**#ROIs**	**Percentage**	**#views**	**#ROIs**
D	DBT	230	1,140	100%					**230**	**1,140**
D	DBT	115	570	50%					115	570
B	MAM	2,242	2,454	100%		230	1,140	100%	2472	3,594
B	MAM	112	123	5%	DBT	115	570	50%	**227**	693
B	MAM	224	245	10%	DBT	115	570	50%	339	815
B	MAM	336	368	15%	DBT	115	570	50%	451	938
B	MAM	448	491	20%	DBT	115	570	50%	563	1,061
B	MAM	561	614	25%	DBT	115	570	50%	676	**1,184**
B	MAM	673	736	30%	DBT	115	570	50%	788	1,306

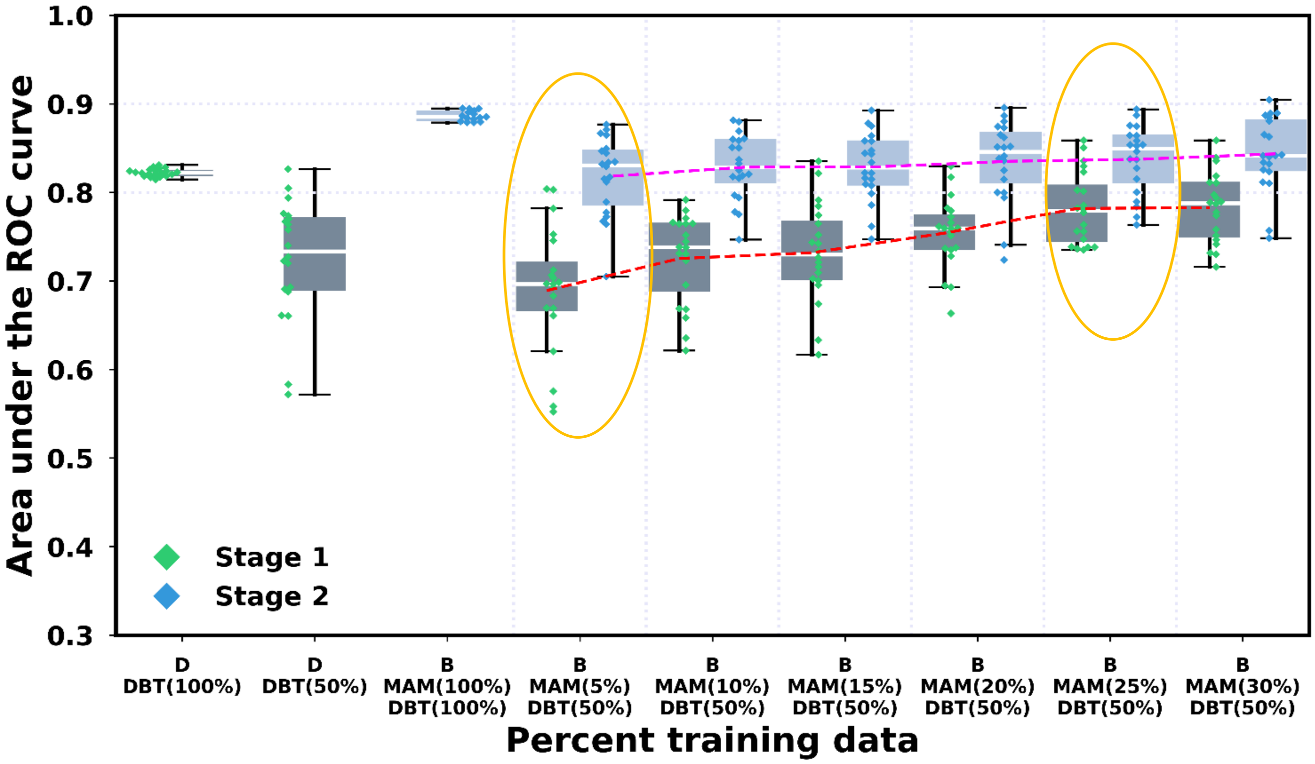

Fig. 10. Box-and-whisker plots of ROI-based AUCs on the DBT test set after various single-stage and multi-stage transfer learning approaches at different mammography and DBT training data sizes as specified along the horizontal axis. Within each pair of dotted vertical lines, the stage 1 AUCs show the performance after training with stage 1 mammography data and the stage 2 AUCs indicate the improvement after the stage 2 DBT transfer learning. The red and pink curves link the mean AUC from the 20 random samplings at each condition to facilitate comparison. The orange ellipses mark the two highlighted conditions shown in Table VI.

VIII. Visualization of deep features from Single- and Multi-stage transfer learning

To visualize the feature spaces of the single-stage and multi-stage transfer learning, we used locally linear embedding (LLE) and *t*-distributed stochastic neighbor embedding (*t*-SNE) to map the features in the four fully connected layers in the DCNN from the high-dimensional space to a two-dimensional (2D) space. Fig. 11 shows the embedded 2D feature spaces for the training and test DBT data.

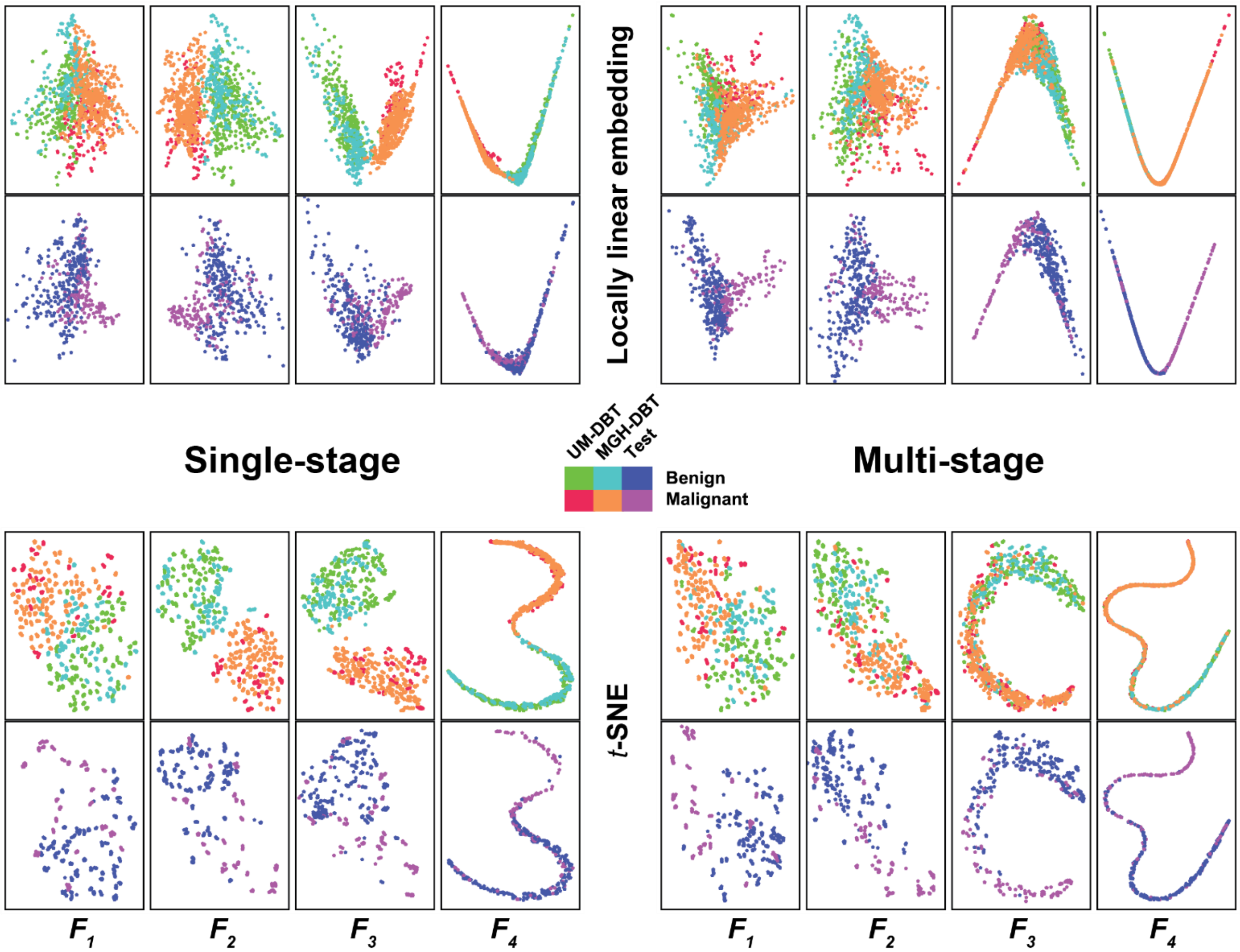

Fig. 11. Two-dimensional locally linear embedding (LLE) and *t*-SNE maps of the training and test DBT samples obtained from the single-stage (trained with 100% DBT) and multi-stage (trained with 100% mammography, 100% DBT) approaches for transfer networks at four fully connected layers. The legend indicates malignant and benign classes for the training and test DBT data.

IX. Activation maps of DBT samples

In all four schemes listed in fig. 2, the CNN was pre-trained with ImageNet data. During transfer learning, the first convolutional layer (C_1) was frozen in all schemes so that they were all the same as that trained with ImageNet data alone. The second convolutional layer (C_2) in both schemes A and C was fined-tuned using mammography data alone because C_2 was frozen during stage 2 training in scheme C. In scheme B, C_2 was fine-tuned with mammography data in stage 1 and DBT data in stage 2. In scheme D, C_2 was directly fine-tuned with DBT data alone. In fig. 12, select activation maps for four DBT ROIs from the training set are shown as examples for convolutional layers C_1 and C_2. The deeper convolutional layers are too small and not shown.

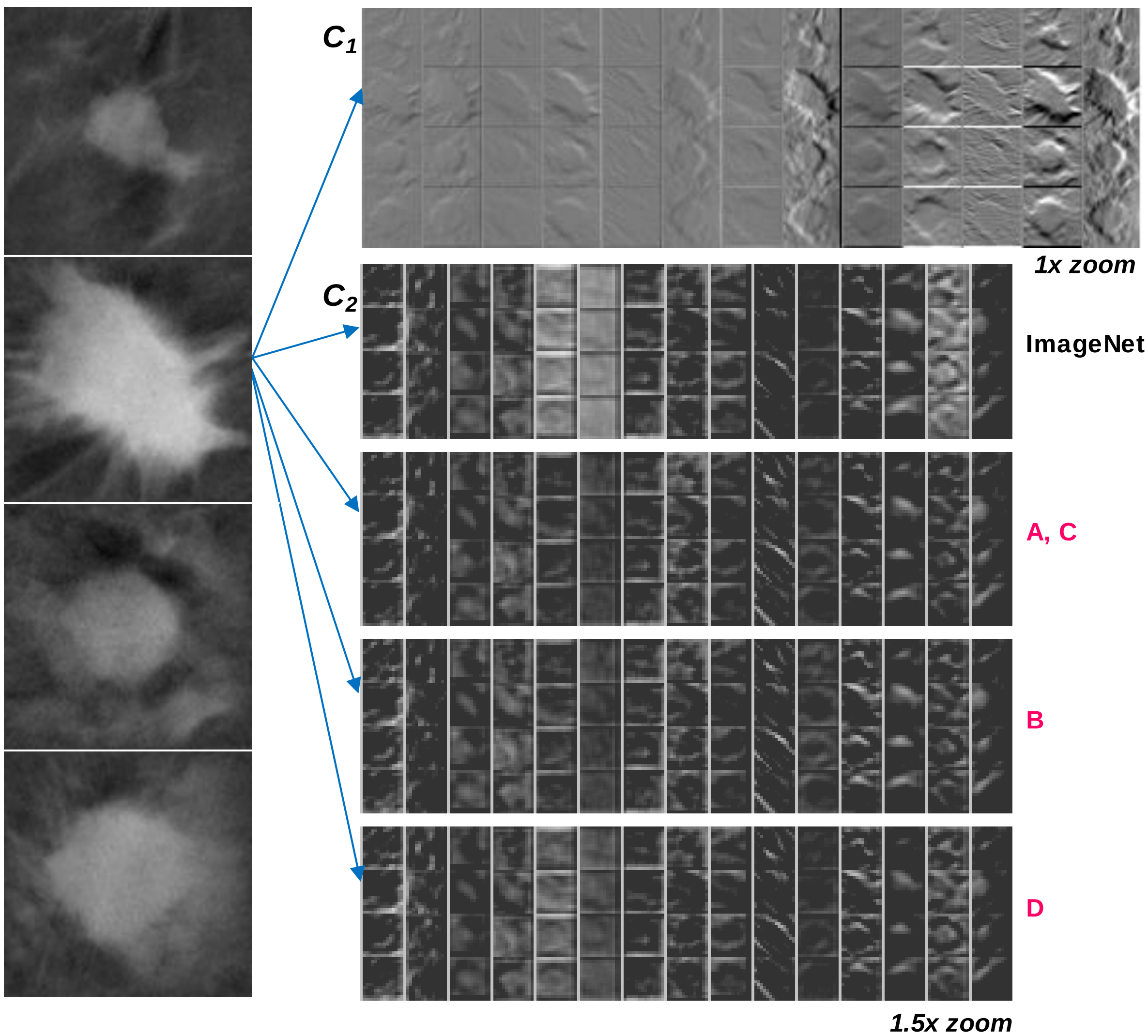

Fig. 12. Examples from the DBT training ROI set are shown on the left. The corresponding activation maps from the first and second convolutional layers are shown on the right. C_1 was frozen in all schemes so that they were identical and only one is shown. C_2 in scheme C was frozen so that it was the same as C_2 from stage 1, which in turn was the same as the C_2 in scheme A.