Appendix

Multi-task Graph Convolutional Neural Network for Calcification Morphology and Distribution Analysis in Mammograms

Fold 1			
Characteristics	**Group Frequency (percentage %)**		**P-value**
	Train n=152	**Test n=43**	
Age, mean (SD)	54.1 (11.5)	56.2 (13.5)	0.350
Mammographic purpose			0.280
Diagnosis	113	32	
Screening	39	11	
Symptoms			0.597
Mass	71 (46.7)	27 (62.8)	
Breast cancer initiate/subsequent screening	39 (25.7)	9 (20.9)	
Lump	21 (13.8)	3 (7.0)	
Positive cases from prior/outside mammography	14 (9.2)	2 (4.7)	
Bloody discharge	2 (1.3)		
Positive cases from prior/outside mammography and abnormal ultrasound finding	1 (0.7)	1 (2.3)	
From abnormal ultrasound	1 (0.7)	1 (2.3)	
Mass + bloody discharge	1 (0.7)		
Lump + bloody discharge	1 (0.7)		
Discharge	1 (0.7)		
Breast density composition			0.845
Fatty breast	4 (2.6)	1 (2.3)	
Scattered fibroglandular density	28 (18.4)	6 (14.0)	
Heterogeneously dense	106 (69.7)	33 (76.7)	
Extremely dense	14 (9.2)	3 (7.0)	
BI-RADS Category assessment			0.234
4A	6 (3.9)	2 (4.7)	
4B	30 (19.7)	3 (7.0)	
4C	30 (19.7)	8 (18.6)	
5	86 (56.6)	30 (69.8)	
Histopathology			0.892
IDC	88 (57.9)	29 (67.4)	
DCIS	11 (25.6)	52 (34.3)	
IDC+DCIS	6 (4.0)	1 (2.3)	
ILC	1 (0.7)	1 (2.3)	
Cellular changes conclusive for malignancy	1 (0.7)		
Adenocarcinoma	1 (0.7)	1 (2.3)	
ILC + IDC + DCIS	1 (0.7)		
Invasive mucinous carcinoma	1 (0.7)		
Distribution descriptors			0.685
Cluster	74 (48.7)	23 (53.5)	
Regional	45 (29.6)	11 (25.6)	
Segmental	23 (15.1)	4 (9.3)	
Diffuse	8 (5.3)	4 (9.3)	
Linear	2 (1.3)	1 (2.3)	
Morphology descriptors			0.842
Pleomorphic	70 (46.1)	18 (41.9)	
Heterogeneous	26 (17.1)	9 (20.9)	
Linear	11 (7.2)	5 (11.6)	
Pleomorphic + amorphous	11 (7.2)	3 (7.0)	
Pleomorphic + linear	10 (6.6)	2 (4.7)	
Amorphous + linear	6 (3.9)	2 (4.7)	
Amorphous + heterogeneous	5 (3.3)	3 (7.0)	
Amorphous	7 (4.6)		
Heterogeneous + linear	5 (3.3)	1 (2.3)	
Heterogeneous + amorphous + linear	1 (0.7)		

Table 1: Basic characteristics of training and testing sets for fold 1 in 5-fold cross validation

Fold 2			
Characteristics	**Group Frequency (percentage %)**		**P-value**
	Train n=154	**Test n=41**	
Age, mean (SD)	55.2 (11.8)	51.9 (12.3)	0.126
Mammographic purpose			0.898
Diagnosis	113 (73.3)	32 (78.1)	
Screening	41 (26.6)	9 (22.0)	
Symptoms			0.786
Mass	79 (51.3)	19 (46.3)	
Breast cancer initiate/subsequent screening	39 (25.3)	9 (20.9)	
Lump	16 (10.4)	8 (19.5)	
Positive cases from prior/outside mammography	12 (7.8)	4 (9.8)	
Bloody discharge	1 (0.6)	1 (2.4)	
Positive cases from prior/outside mammography and abnormal ultrasound finding	2 (1.3)		
From abnormal ultrasound	2 (1.3)		
Mass + bloody discharge	1 (0.6)		
Lump + bloody discharge	1 (0.6)		
Discharge	1 (0.6)		
Breast density composition			0.417
Fatty breast	5 (3.2)		
Scattered fibroglandular density	25 (16.2)	9 (22.0)	
Heterogeneously dense	109 (70.8)	30 (73.2)	
Extremely dense	15 (9.7)	2 (4.9)	
BI-RADS Category assessment			0.864
4A	6 (3.9)	2 (4.9)	
4B	25 (16.2)	8 (19.5)	
4C	29 (18.8)	9 (22.0)	
5	94 (61.0)	22 (53.7)	
Histopathology			0.535
IDC	92 (59.7)	25 (60.9)	
DCIS	49 (31.8)	14 (34.1)	
IDC+DCIS	5 (3.2)	2 (4.8)	
ILC	2 (1.3)		
Cellular changes conclusive for malignancy	1 (0.6)		
Adenocarcinoma	2 (1.3)		
ILC + IDC + DCIS	1 (0.6)		
Invasive mucinous carcinoma	1 (0.6)		
Distribution descriptors			0.794
Cluster	74 (48.1)	23 (56.1)	
Regional	47 (30.5)	9 (22.0)	
Segmental	21 (13.6)	6 (14.6)	
Diffuse	10 (6.5)	2 (4.9)	
Linear	2 (1.3)	1 (2.4)	
Morphology descriptors			0.578
Pleomorphic	68 (44.2)	20 (48.8)	
Heterogeneous	29 (18.8)	6 (14.6)	
Linear	12 (7.8)	4 (9.8)	
Pleomorphic + amorphous	10 (6.5)	4 (9.8)	
Pleomorphic + linear	9 (5.8)	3 (7.3)	
Amorphous + linear	7 (4.5)	1 (2.4)	
Amorphous + heterogeneous	8 (5.2)		
Amorphous	4 (2.6)	3 (7.3)	
Heterogeneous + linear	6 (3.9)		
Heterogeneous + amorphous + linear	1 (0.6)		

Table 2: Basic characteristics of training and testing sets for fold 2 in 5-fold cross validation

Fold 3			
Characteristics	**Group Frequency (percentage %)**		**P-value**
	Train n=152	**Test n=43**	
Age, mean (SD)	54.3 (12.0)	55.2 (11.9)	0.694
Mammographic purpose			0.441
Diagnosis	110 (71.9)	35 (83.3)	
Screening	43 (28.2)	7 (16.7)	
Symptoms			0.231
Mass	76 (49.7)	22 (52.4)	
Breast cancer initiate/subsequent screening	41 (26.8)	7 (16.7)	
Lump	18 (11.8)	6 (14.3)	
Positive cases from prior/outside mammography	11 (7.2)	5 (11.9)	
Bloody discharge	2 (1.3)		
Positive cases from prior/outside mammography and abnormal ultrasound finding	2 (1.3)		
From abnormal ultrasound	2 (1.3)		
Mass + bloody discharge	1 (0.7)		
Lump + bloody discharge		1 (2.4)	
Discharge		1 (2.4)	
Breast density composition			0.073
Fatty breast	2 (1.3)	3 (7.1)	
Scattered fibroglandular density	28 (18.3)	6 (14.3)	
Heterogeneously dense	112 (73.2)	27 (64.3)	
Extremely dense	11 (7.2)	6 (14.3)	
BI-RADS Category assessment			0.846
4A	7 (4.6)	1 (2.4)	
4B	26 (17.0)	7 (16.7)	
4C	31 (20.3)	7 (16.7)	
5	89 (58.2)	27 (64.3)	
Histopathology			0.921
IDC	89 (58.2)	28 (66.7)	
DCIS	52 (34.0)	11 (26.2)	
IDC+DCIS	5 (3.4)	2 (4.8)	
ILC	1 (0.7)	1 (2.4)	
Cellular changes conclusive for malignancy	1 (0.7)		
Adenocarcinoma	2 (1.3)		
ILC + IDC + DCIS	1 (0.7)		
Invasive mucinous carcinoma	1 (0.7)		
Distribution descriptors			0.870
Cluster	79 (51.6)	18 (42.9)	
Regional	42 (27.5)	14 (33.3)	
Segmental	21 (13.7)	6 (14.3)	
Diffuse	9 (5.9)	3 (7.1)	
Linear	2 (1.3)	1 (2.4)	
Morphology descriptors			0.429
Pleomorphic	64 (41.8)	24 (57.1)	
Heterogeneous	29 (19.0)	6 (14.3)	
Linear	13 (8.5)	3 (7.1)	
Pleomorphic + amorphous	13 (8.5)	1 (2.4)	
Pleomorphic + linear	11 (7.2)	1 (2.4)	
Amorphous + linear	6 (3.9)	2 (4.8)	
Amorphous + heterogeneous	5 (3.3)	3 (7.1)	
Amorphous	7 (4.6)		
Heterogeneous + linear	4 (2.6)	2 (4.8)	
Heterogeneous + amorphous + linear	1 (0.7)		

Table 3: Basic characteristics of training and testing sets for fold 3 in 5-fold cross validation

Fold 4			
Characteristics	**Group**		**P-value**
	Frequency (percentage %)		
	Train n=153	**Test n=42**	
Age, mean (SD)	54.3 (11.8)	55.2 (12.6)	0.707
Mammographic purpose			0.878
Diagnosis	115 (75.2)	30 (71.4)	
Screening	38 (24.9)	12 (28.6)	
Symptoms			0.585
Mass	75 (49.0)	23 (54.8)	
Breast cancer initiate/subsequent screening	36 (23.5)	12 (28.6)	
Lump	20 (13.1)	4 (9.5)	
Positive cases from prior/outside mammography	14 (9.2)	2 (4.8)	
Bloody discharge	2 (1.3)		
Positive cases from prior/outside mammography and abnormal ultrasound finding	2 (1.3)		
From abnormal ultrasound	2 (1.3)		
Mass + bloody discharge		1 (2.4)	
Lump + bloody discharge	1 (0.7)		
Discharge	1 (0.7)		
Breast density composition			0.323
Fatty breast	4 (2.6)	1 (2.4)	
Scattered fibroglandular density	28 (18.3)	6 (14.3)	
Heterogeneously dense	105 (68.6)	34 (81.0)	
Extremely dense	16 (10.5)	1 (2.4)	
BI-RADS Category assessment			0.573
4A	7 (4.6)	1 (2.4)	
4B	24 (15.7)	9 (21.4)	
4C	28 (18.3)	10 (23.8)	
5	94 (61.4)	22 (52.4)	
Histopathology			0.764
IDC	89 (58.2)	28 (66.7)	
DCIS	52 (34.0)	11 (26.2)	
IDC+DCIS	6 (4.0)	1 (2.4)	
ILC	2 (1.3)		
Cellular changes conclusive for malignancy	1 (0.7)		
Adenocarcinoma	1 (0.7)	1 (2.4)	
ILC + IDC + DCIS	1 (0.7)		
Invasive mucinous carcinoma		1 (2.4)	
Distribution descriptors			0.178
Cluster	78 (51.0)	19 (45.2)	
Regional	41 (26.8)	15 (35.7)	
Segmental	19 (12.4)	8 (19.0)	
Diffuse	12 (7.8)		
Linear	3 (2.0)		
Morphology descriptors			0.922
Pleomorphic	71 (46.4)	17 (40.5)	
Heterogeneous	25 (16.3)	10 (23.8)	
Linear	12 (7.8)	4 (9.5)	
Pleomorphic + amorphous	12 (7.8)	2 (4.8)	
Pleomorphic + linear	8 (5.2)	4 (9.5)	
Amorphous + linear	6 (3.9)	2 (4.8)	
Amorphous + heterogeneous	7 (4.6)	1 (2.4)	
Amorphous	6 (3.9)	1 (2.4)	
Heterogeneous + linear	5 (3.3)	1 (2.4)	
Heterogeneous + amorphous + linear	1 (0.7)		

Table 4: Basic characteristics of training and testing sets for fold 4 in 5-fold cross validation

Fold 5			
Characteristics	**Group**		**P-value**
	Frequency (percentage %)		
	Train n=154	**Test n=41**	
Age, mean (SD)	55.1 (11.8)	52.4 (12.4)	0.215
Mammographic purpose			0.411
Diagnosis	115 (74.7)	30 (73.1)	
Screening	39 (25.3)	11 (26.8)	
Symptoms			0.765
Mass	81 (52.6)	17 (41.5)	
Breast cancer initiate/subsequent screening	37 (24.0)	11 (26.8)	
Lump	17 (11.0)	7 (17.1)	
Positive cases from prior/outside mammography	13 (8.4)	3 (7.3)	
Bloody discharge	1 (0.6)	1 (2.4)	
Positive cases from prior/outside mammography and abnormal ultrasound finding	1 (0.6)	1 (2.4)	
From abnormal ultrasound	1 (0.6)	1 (2.4)	
Mass + bloody discharge	1 (0.6)		
Lump + bloody discharge	1 (0.6)		
Discharge	1 (0.6)		
Breast density composition			0.523
Fatty breast	5 (3.2)		
Scattered fibroglandular density	26 (16.9)	8 (19.5)	
Heterogeneously dense	111 (72.1)	28 (68.3)	
Extremely dense	12 (7.8)	5 (12.2)	
BI-RADS Category assessment			0.157
4A	6 (3.9)	2 (4.9)	
4B	27 (17.5)	6 (14.6)	
4C	25 (16.2)	13 (31.7)	
5	96 (62.3)	20 (48.8)	
Histopathology			0.061
IDC	96 (62.3)	21 (51.2)	
DCIS	47 (30.5)	16 (39.0)	
IDC+DCIS	6 (3.8)	1 (2.4)	
ILC	2 (1.3)		
Cellular changes conclusive for malignancy		1 (2.4)	
Adenocarcinoma	2 (1.3)		
ILC + IDC + DCIS		1 (2.4)	
Invasive mucinous carcinoma	1 (0.6)		
Distribution descriptors			0.288
Cluster	74 (48.1)	23 (56.1)	
Regional	49 (31.8)	7 (17.1)	
Segmental	19 (12.3)	8 (19.5)	
Diffuse	9 (5.8)	3 (7.3)	
Linear	3 (1.9)		
Morphology descriptors			0.371
Pleomorphic	71 (46.1)	17 (41.5)	
Heterogeneous	30 (19.5)	5 (12.2)	
Linear	11 (7.1)	5 (12.2)	
Pleomorphic + amorphous	10 (6.5)	4 (9.8)	
Pleomorphic + linear	10 (6.5)	2 (4.9)	
Amorphous + linear	7 (4.5)	1 (2.4)	
Amorphous + heterogeneous	7 (4.5)	1 (2.4)	
Amorphous	4 (2.6)	3 (7.3)	
Heterogeneous + linear	4 (2.6)	2 (4.9)	
Heterogeneous + amorphous + linear		1 (2.4)	

Table 5: Basic characteristics of training and testing sets for fold 5 in 5-fold cross validation

Type	Methods	ROC-AUC of distribution classes±std				
		Diffuse	**Regional**	**Cluster**	**Linear**	**Segmental**
Baselines	**ResNet**	0.766±0.107	0.672±0.094	0.640±0.077	0.638±0.109	0.582±0.096
	DenseNet	0.638±0.109	0.648±0.091	0.628±0.082	0.476±0.375	0.482±0.060
	MobileNet	0.660±0.147	0.634±0.074	0.624±0.033	0.864±0.099	0.654±0.109
	EfficientNet	0.670±0.148	0.720±0.094	0.624±0.110	0.802±0.191	0.556±0.079
Proposed	**Multi-task, multi-graph, 8-layer GCN**	**0.944±0.059**	**0.754±0.058**	**0.798±0.055**	**0.544±0.060**	**0.680±0.078**

(a) ROC-AUC performance on each type of dsitribution descriptor class

Type	Methods	ROC-AUC of morphology classes±std			
		Coarse heterogeneous	**Amorphous**	**Fine pleomorphic**	**Fine linear (fine-linear branching)**
Baselines	**ResNet**	0.528±0.045	0.578±0.025	0.548±0.027	0.554±0.074
	DenseNet	0.546±0.073	0.634±0.077	0.592±0.040	0.548±0.079
	MobileNet	0.572±0.034	0.564±0.036	0.556±0.005	0.546±0.057
	EfficientNet	0.548±0.016	0.582±0.036	0.558±0.032	0.540±0.042
Proposed	**Multi-task, multi-graph, 8-layer GCN**	**0.758±0.053**	**0.620±0.088**	**0.614±0.033**	**0.616±0.090**

(b) ROC-AUC performance on each type of morphology descriptor class

Table 6: The ROC-AUC performance comparison on TMU dataset between baseline models and proposed model on distribution and morphology classification. ROC-AUC is evaluated on each type of distribution and morphology descriptor.

Type	Methods	ROC-AUC of distribution classes±std				
		Diffuse	**Regional**	**Cluster**	**Linear**	**Segmental**
Baselines	**ResNet**	0.900±0.052	0.700±0.148	0.546±0.057	0.460±0.075	0.554±0.066
	DenseNet	0.958±0.031	0.596±0.183	0.638±0.073	0.614±0.052	0.528±0.111
	MobileNet	0.784±0.143	0.792±0.055	0.640±0.042	0.490±0.046	0.600±0.030
	EfficientNet	0.884±0.152	0.640±0.092	0.580±0.079	0.628±0.091	0.620±0.036
Proposed	**Multi-task, multi-graph, 8-layer GCN**	**0.984±0.017**	**0.904±0.053**	**0.870±0.018**	**0.712±0.070**	**0.896±0.036**

(a) ROC-AUC performance on each type of morphology descriptor class

Type	Methods	ROC-AUC of morphology classes±std				
		Amorphous	**Fine pleomorphic**	**Fine linear (fine-linear branching)**	**Punctate**	**Round and regular**
Baselines	**ResNet**	0.568±0.031	0.538±0.050	0.566±0.054	0.572±0.029	0.776±0.042
	DenseNet	0.544±0.037	0.522±0.065	0.556±0.050	0.600±0.087	0.832±0.081
	MobileNet	0.572±0.056	0.550±0.039	0.506±0.072	0.554±0.049	0.816±0.075
	EfficientNet	0.564±0.037	0.546±0.048	0.576±0.080	0.556±0.043	0.822±0.101
Proposed	**Multi-task, multi-graph, 8-layer GCN**	**0.586±0.125**	**0.610±0.116**	**0.664±0.032**	**0.740±0.070**	**0.906±0.122**

(b) ROC-AUC performance on each type of morphology descriptor class

Table 7: The ROC-AUC performance comparison on CBIS-DDSM dataset between baseline models and proposed model on distribution and morphology classification. ROC-AUC is evaluated on each type of distribution and morphology descriptor.