A Real Use Case of Semi-Supervised Learning for Mammogram Classification in a Local Clinic of Costa Rica — Supplementary Material

Saúl Calderón-Ramírez · Diego Murillo-Hernández · Kevin Rojas-Salazar · David Elizondo · Shengxiang Yang · Armaghan Moemeni · Miguel A. Molina-Cabello

This document consists of some supplementary experimental results to the paper "Real Use Case of Semi-Supervised Learning for Mammogram Classification in a Local Clinic of Costa Rica", published in the journal Medical & Biological Engineering & Computing.

Saúl Calderón-Ramírez · David Elizondo · Shengxiang Yang
Centre for Computational Intelligence (CCI), De Montfort University, United Kingdom

Saúl Calderón-Ramírez · Diego Murillo-Hernández · Kevin Rojas-Salazar
PARMA Research Group, Instituto Tecnológico de Costa Rica, Costa Rica

Diego Murillo-Hernández · Kevin Rojas-Salazar
School of Computing Engineering, Instituto Tecnológico de Costa Rica, Costa Rica

Armaghan Moemeni
School of Computer Science, University of Nottingham, United Kingdom

Miguel A. Molina-Cabello
Department of Computer Languages and Computer Science. University of Málaga, Spain
Instituto de Investigación Biomédica de Málaga - IBIMA, Spain

E-mail: sacalderon@itcr.ac.cr, diemurillo@ic-itcr.ac.cr, kevin.rojas7@estudiantec.cr, elizondo@dmu.ac.uk, syang@dmu.ac.uk, armaghan.moemeni@nottingham.ac.uk, miguelangel@lcc.uma.es

Table 1: Results of Configurations **SSDL+FT** and **S+FT**, using **INbreast** as source dataset with additional model architectures

(a) ResNet-152

n_t^l	Metric	SSDL		Supervised	
		$\bar{x}$	s	$\bar{x}$	s
20	G-Mean	**0.6774**	0.1167	0.6767	0.1021
	F2-Score	**0.3577**	0.1349	0.3155	0.0828
	Accuracy	**0.6851**	0.1903	0.5978	0.1021
	Recall	0.7250	0.2189	**0.7917**	0.1632
	Specificity	**0.6838**	0.2064	0.5883	0.1062
	Precision*	**0.1407**	0.0869	0.0946	0.0300
40	G-Mean	**0.6762**	0.1170	0.6705	0.1037
	F2-Score	**0.3422**	0.1285	0.3099	0.0814
	Accuracy	**0.7075**	0.1237	0.6443	0.0940
	Recall	0.6750	0.2058	**0.7250**	0.2189
	Specificity	**0.7080**	0.1341	0.6409	0.1017
	Precision	**0.1230**	0.0640	0.0975	0.0283
60	G-Mean	0.6539	0.1384	**0.6774**	0.0787
	F2-Score	0.3226	0.1386	**0.3388**	0.0973
	Accuracy	0.6590	0.1547	**0.7034**	0.1353
	Recall	0.6667	0.2041	**0.6667**	0.1179
	Specificity	0.6584	0.1599	**0.7051**	0.1442
	Precision	0.1201	0.0803	**0.1231**	0.0567

(b) EfficientNet-b0

n_t^l	Metric	SSDL		Supervised	
		$\bar{x}$	s	$\bar{x}$	s
20	G-Mean	**0.6512**	0.1081	0.6393	0.0603
	F2-Score	**0.3282**	0.1142	0.2954	0.0538
	Accuracy	**0.7440**	0.1433	0.7011	0.1072
	Recall	**0.6000**	0.2108	0.5917	0.1208
	Specificity	**0.7519**	0.1565	0.7062	0.1157
	Precision	**0.1309**	0.0624	0.1041	0.0338
40	G-Mean	**0.6752**	0.0977	0.6401	0.0754
	F2-Score	**0.3331**	0.0999	0.2920	0.0598
	Accuracy	0.7090	0.1174	**0.7139**	0.0768
	Recall	**0.6667**	0.2041	0.5750	0.1208
	Specificity	0.7107	0.1273	**0.7209**	0.0797
	Precision	**0.1161**	0.0470	0.0999	0.0232
60	G-Mean	**0.6620**	0.0919	0.6422	0.0869
	F2-Score	**0.3274**	0.1031	0.2975	0.0808
	Accuracy	**0.7229**	0.1279	0.7161	0.0836
	Recall	**0.6250**	0.1768	0.5750	0.1208
	Specificity	**0.7282**	0.1376	0.7234	0.0851
	Precision	**0.1196**	0.0540	0.1036	0.0352

Table 2: Results of Configurations **SSDL+FT** and **S+FT**, using **Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM)** as source dataset with additional model architectures

(a) ResNet-152

n_t^l	Metric	SSDL		Supervised	
		$\bar{x}$	s	$\bar{x}$	s
20	G-Mean*	**0.6575**	0.1075	0.5857	0.0598
	F2-Score	**0.3068**	0.1151	0.2527	0.0672
	Accuracy	0.5910	0.1184	**0.6828**	0.1286
	Recall*	**0.7583**	0.1687	0.5000	0.0000
	Specificity*	0.5825	0.1234	**0.6925**	0.1354
	Precision	**0.0959**	0.0538	0.0917	0.0432
40	G-Mean	**0.5947**	0.1502	0.5759	0.0626
	F2-Score	**0.2610**	0.1228	0.2426	0.0643
	Accuracy	0.6541	0.0909	**0.6616**	0.1327
	Recall	**0.5833**	0.2966	0.5000	0.0000
	Specificity	0.6569	0.1002	**0.6703**	0.1397
	Precision	0.0839	0.0416	**0.0848**	0.0368
60	G-Mean	**0.6153**	0.1250	0.5561	0.0769
	F2-Score	**0.2706**	0.1003	0.2207	0.0642
	Accuracy	**0.6869**	0.0613	0.6463	0.0909
	Recall	**0.5750**	0.2372	0.4750	0.0791
	Specificity	**0.6931**	0.0696	0.6554	0.0940
	Precision	**0.0907**	0.0399	0.0731	0.0306

(b) EfficientNet-b0

n_t^l	Metric	SSDL		Supervised	
		$\bar{x}$	s	$\bar{x}$	s
20	G-Mean	**0.5982**	0.0753	0.5824	0.0489
	F2-Score	**0.2521**	0.0550	0.2467	0.0475
	Accuracy	**0.6537**	0.0938	0.5886	0.1426
	Recall	0.5750	0.2058	**0.6083**	0.1419
	Specificity	**0.6579**	0.1060	0.5866	0.1558
	Precision	**0.0805**	0.0181	0.0757	0.0192
40	G-Mean	0.6055	0.1048	**0.6130**	0.0519
	F2-Score	0.2607	0.0853	**0.2884**	0.0668
	Accuracy	0.6693	0.0894	**0.7182**	0.1443
	Recall	**0.5750**	0.2372	0.5333	0.1054
	Specificity	0.6742	0.0993	**0.7264**	0.1562
	Precision	0.0839	0.0262	**0.1102**	0.0436
60	G-Mean	**0.6014**	0.0971	0.6013	0.0618
	F2-Score	0.2612	0.0835	**0.2792**	0.0882
	Accuracy	0.6810	0.1020	**0.7187**	0.1391
	Recall	**0.5500**	0.1972	0.5000	0.0000
	Specificity	0.6878	0.1106	**0.7300**	0.1459
	Precision	0.0890	0.0362	**0.1200**	0.0869

---

*Statistic significance (p-values $<$ 0.05) for average differences between results of Semi-supervised Deep Learning (SSDL) and supervised models