

**Supplementary Material of the
WACV 2025 Paper:
Remote Blood Pressure Estimation from
Facial Videos using Transfer Learning:
Leveraging PPG to rPPG Conversion**

1 Performance Comparison of Blood Pressure Estimation with PPG Data

Dataset	Models	Normalization	Task-based Learning	with HR		without HR	
				MAE (mmHg) SBP	MAE (mmHg) DBP	MAE (mmHg) SBP	MAE (mmHg) DBP
MIMIC-III from Kaggle	Alexnet	Per Signal	Multi-task	12.020	5.752	12.165	5.824
	ResNet			12.284	6.000	12.640	6.317
	BiLSTM			11.857	5.943	11.121	5.523
	UniLSTM			18.907	8.233	19.210	8.309
	Slapničar et al.			14.152	7.662	12.882	7.333
	BP-Net			12.010	5.846	11.980	5.887
MIMIC-III from Kaggle	Alexnet	Per Signal	Single-task	12.397	5.689	12.227	5.822
	ResNet			11.787	6.247	11.517	5.701
	BiLSTM			12.011	5.687	11.037	6.761
	UniLSTM			18.950	6.714	19.230	6.552
	Slapničar et al.			13.867	6.383	12.874	6.318
	BP-Net			12.045	5.814	12.046	5.831
MIMIC-III from Kaggle	Alexnet	Per Dataset	Multi-task	11.913	5.684	12.128	5.816
	ResNet			12.651	6.125	12.823	6.010
	BiLSTM			11.404	5.503	11.321	5.561
	UniLSTM			12.481	6.213	13.346	6.627
	Slapničar et al.			13.883	7.594	11.980	7.286
	BP-Net			11.367	5.667	11.353	5.602
MIMIC-III from Kaggle	Alexnet	Per Dataset	Single-task	12.195	5.903	12.206	5.650
	ResNet			13.203	6.126	12.212	6.148
	BiLSTM			12.881	7.146	13.032	7.729
	UniLSTM			13.169	7.666	13.389	7.783
	Slapničar et al.			14.625	5.804	11.888	5.958
	BP-Net			11.333	5.587	11.446	5.586
MIMIC-III from PulseDB	Alexnet	Per Signal	Multi-task	15.229	8.798	15.132	8.573
	ResNet			14.907	8.410	14.974	8.249
	BiLSTM			15.420	8.451	14.810	8.545
	UniLSTM			17.720	9.572	17.046	8.458
	Slapničar et al.			15.606	8.963	16.105	9.045
	BP-Net			17.203	8.781	17.141	8.778
MIMIC-III from PulseDB	Alexnet	Per Signal	Single-task	15.427	8.728	15.187	8.502
	ResNet			14.928	8.409	15.078	8.417
	BiLSTM			14.907	8.445	15.074	8.179
	UniLSTM			17.112	8.436	17.003	8.294
	Slapničar et al.			16.015	8.619	14.936	8.179
	BP-Net			17.225	8.752	17.170	8.753
MIMIC-III from PulseDB	Alexnet	Per Dataset	Multi-task	15.138	8.540	15.685	8.539
	ResNet			14.965	8.429	15.115	8.414
	BiLSTM			15.199	8.803	14.860	8.518
	UniLSTM			17.667	9.520	15.155	8.574
	Slapničar et al.			15.706	9.336	15.188	8.435
	BP-Net			17.429	8.691	17.394	8.711
MIMIC-III from PulseDB	Alexnet	Per Dataset	Single-task	15.288	8.573	15.218	8.586
	ResNet			14.831	8.550	15.250	8.487
	BiLSTM			14.714	9.497	14.887	8.419
	UniLSTM			15.915	9.466	14.893	9.603
	Slapničar et al.			15.952	8.737	16.168	8.315
	BP-Net			17.504	8.680	17.384	8.715

Table 1: Performance comparison of deep learning models trained and tested on different PPG datasets. Mean absolute error (MAE) in mmHg was used as an evaluation metric. The best performing models for each dataset were in bold.

2 Performance Comparison of Blood Pressure Estimation by Transfer Learning with rPPG Data

Pretrain Dataset	Models	Normalization	Loss Function	with HR		without HR	
				MAE (mmHg) SBP	DBP	MAE (mmHg) SBP	DBP
MIMIC-III from Kaggle	Alexnet	Per Signal	Mean Absolute Error	14.462	10.158	15.907	11.166
	ResNet			10.270	10.435	12.012	9.856
	BiLSTM			9.886	9.381	12.081	10.083
	Slapničar et al. BP-Net			10.711	10.084	10.803	10.190
				11.353	10.066	11.136	9.602
MIMIC-III from Kaggle	Alexnet	Per Signal	Mean Squared Error	13.784	9.768	15.507	11.063
	ResNet			10.058	10.916	11.790	10.263
	BiLSTM			9.927	9.986	11.460	10.516
	Slapničar et al. BP-Net			10.879	10.828	11.136	10.037
				11.338	9.943	10.616	10.135
MIMIC-III from Kaggle	Alexnet	Per Dataset	Mean Absolute Error	13.200	9.364	14.995	9.559
	ResNet			9.731	10.192	11.220	10.224
	BiLSTM			8.721	8.653	12.053	9.799
	Slapničar et al. BP-Net			13.587	11.267	10.421	10.226
				11.422	10.181	11.601	10.107
MIMIC-III from Kaggle	Alexnet	Per Dataset	Mean Squared Error	12.163	9.029	13.234	10.639
	ResNet			10.159	10.173	11.054	10.527
	BiLSTM			9.056	9.424	12.031	10.698
	Slapničar et al. BP-Net			14.033	10.700	10.587	10.599
				11.154	9.693	11.846	10.104
MIMIC-III from PulseDB	Alexnet	Per Signal	Mean Absolute Error	14.607	9.130	14.331	9.371
	ResNet			10.550	10.271	11.692	10.289
	BiLSTM			9.644	9.773	11.529	10.355
	Slapničar et al. BP-Net			11.047	10.425	10.846	10.285
				11.271	10.247	11.209	9.985
MIMIC-III from PulseDB	Alexnet	Per Signal	Mean Squared Error	13.752	9.374	14.764	9.197
	ResNet			10.982	10.401	11.419	10.232
	BiLSTM			9.443	10.025	11.637	10.362
	Slapničar et al. BP-Net			11.837	10.535	11.328	9.993
				10.921	9.782	11.584	9.532
MIMIC-III from PulseDB	Alexnet	Per Dataset	Mean Absolute Error	14.822	9.644	14.382	10.237
	ResNet			10.312	10.389	12.731	11.232
	BiLSTM			9.359	10.307	12.824	9.932
	Slapničar et al. BP-Net			13.439	11.299	10.809	10.330
				11.361	9.712	11.664	9.649
MIMIC-III from PulseDB	Alexnet	Per Dataset	Mean Squared Error	13.692	9.794	15.296	9.404
	ResNet			10.502	9.900	12.517	11.184
	BiLSTM			9.625	10.026	12.981	10.468
	Slapničar et al. BP-Net			12.677	10.972	10.709	10.135
				11.590	9.841	11.700	9.739
No Pretrain	Alexnet	Per Signal	Mean Absolute Error	15.344	10.926	15.702	11.410
	ResNet			11.315	11.114	12.716	10.869
	BiLSTM			12.683	10.385	13.645	10.489
	Slapničar et al. BP-Net			16.224	11.565	13.308	11.530
				12.266	10.969	12.399	11.065
No Pretrain	Alexnet	Per Signal	Mean Squared Error	14.656	10.421	17.471	11.357
	ResNet			12.043	11.642	12.318	10.992
	BiLSTM			12.004	10.953	13.576	10.912
	Slapničar et al. BP-Net			16.000	11.754	13.499	11.078
				12.541	11.910	12.489	10.885
No Pretrain	Alexnet	Per Dataset	Mean Absolute Error	15.685	10.745	15.980	10.977
	ResNet			11.243	11.326	13.241	11.692
	BiLSTM			10.603	11.110	14.011	10.515
	Slapničar et al. BP-Net			20.735	12.274	16.146	11.126
				12.241	10.838	12.480	11.113
No Pretrain	Alexnet	Per Dataset	Mean Squared Error	14.353	10.460	16.189	11.575
	ResNet			11.173	11.268	13.452	11.897
	BiLSTM			10.141	10.810	13.592	11.707
	Slapničar et al. BP-Net			20.673	12.450	16.467	11.534
				12.610	10.823	12.556	10.954

Table 2: Performance comparison of deep learning models with different training settings on the rPPG-LUH. The 4-fold averaged results were reported. Mean absolute error (MAE) in mmHg was used as an evaluation metric. The best performing model for rPPG-LUH was in bold.

3 Results of the Statistical Tests of Blood Pressure Estimation by Transfer Learning with rPPG Data

3.1 Results of the paired t-tests based on pretrain setting (pretrained with PPG datasets vs No pretrain)

3.1.1 Pretrained with MIMIC-III from Kaggle vs No pretrain

Alexnet

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	3.499	4.699e-04	-25.849	1.819e-141
Per Signal	Mean Squared Error	56.479	1.374e-47	-23.225	1.853e-115
Per Dataset	Mean Absolute Error	13.956	9.574e-44	18.901	5.882e-78
Per Dataset	Mean Squared Error	-21.060	8.053e-96	-36.696	3.105e-272

Table 3: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from Kaggle vs No pretrain) for SBP prediction with Alexnet.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-11.914	1.914e-32	-35.964	2.479e-262
Per Signal	Mean Squared Error	-10.195	2.940e-24	-11.588	8.398e-31
Per Dataset	Mean Absolute Error	-36.780	2.234e-273	1.986	4.704e-02
Per Dataset	Mean Squared Error	-13.171	3.346e-39	-39.873	0.000

Table 4: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from Kaggle vs No pretrain) for DBP prediction with Alexnet.

Resnet

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-30.869	1.398e-197	51.076	0.000
Per Signal	Mean Squared Error	36.178	3.279e-265	-38.575	2.869e-298
Per Dataset	Mean Absolute Error	-42.498	0.000	-23.001	2.380e-113
Per Dataset	Mean Squared Error	-13.094	9.065e-39	-9.989	2.335e-23

Table 5: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from Kaggle vs No pretrain) for SBP prediction with ResNet.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	43.635	0.000	15.659	1.925e-54
Per Signal	Mean Squared Error	-44.913	0.000	-43.029	0.000
Per Dataset	Mean Absolute Error	-9.579	1.280e-21	-31.179	2.600e-201
Per Dataset	Mean Squared Error	34.443	3.139e-242	-53.687	0.000

Table 6: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from Kaggle vs No pretrain) for DBP prediction with ResNet.

BiLSTM

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-6.628	3.620e-11	-20.633	3.777e-92
Per Signal	Mean Squared Error	3.517	4.391e-04	8.348	8.106e-17
Per Dataset	Mean Absolute Error	26.756	5.277e-151	22.070	8.880e-105
Per Dataset	Mean Squared Error	-40.819	0.000	-3.009	2.628e-03

Table 7: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from Kaggle vs No pretrain) for SBP prediction with BiLSTM.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-16.316	7.059e-59	-2.908	3.644e-03
Per Signal	Mean Squared Error	16.078	2.989e-57	-11.880	2.824e-32
Per Dataset	Mean Absolute Error	-15.988	1.202e-56	8.932	5.143e-19
Per Dataset	Mean Squared Error	-3.310	9.378e-04	11.786	8.510e-32

Table 8: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from Kaggle vs No pretrain) for DBP prediction with BiLSTM.

Slapničar et al.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-8.593	1.015e-17	23.173	5.718e-115
Per Signal	Mean Squared Error	-2.316	7.522e-03	25.091	1.048e-133
Per Dataset	Mean Absolute Error	11.056	3.327e-28	-13.483	5.582e-41
Per Dataset	Mean Squared Error	42.698	0.000	9.390	7.694e-21

Table 9: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from Kaggle vs No pretrain) for SBP prediction with Slapničar et al..

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-24.614	6.210e-129	45.108	0.000
Per Signal	Mean Squared Error	32.269	1.248e-214	14.625	8.192e-48
Per Dataset	Mean Absolute Error	5.924	3.277e-09	8.619	8.072e-18
Per Dataset	Mean Squared Error	33.102	4.841e-225	21.978	6.071e-104

Table 10: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from Kaggle vs No pretrain) for DBP prediction with Slapničar et al..

BP-Net

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	2.911	3.621e-02	-4.736	2.222e-06
Per Signal	Mean Squared Error	-5.363	8.412e-08	4.557	5.258e-06
Per Dataset	Mean Absolute Error	-2.754	7.949e-03	8.007	1.342e-15
Per Dataset	Mean Squared Error	6.061	1.411e-09	-3.896	9.879e-05

Table 11: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from Kaggle vs No pretrain) for SBP prediction with BP-Net.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	11.211	5.983e-29	-4.647	3.417e-06
Per Signal	Mean Squared Error	7.188	7.167e-13	5.186	2.206e-07
Per Dataset	Mean Absolute Error	2.148	2.508e-02	22.870	4.041e-112
Per Dataset	Mean Squared Error	6.706	2.145e-11	2.584	1.133e-02

Table 12: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from Kaggle vs No pretrain) for DBP prediction with BP-Net.

3.1.2 Pretrained with MIMIC-III from PulseDB vs No pretrain

Alexnet

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	8.301	1.201e-16	-23.470	8.722e-118
Per Signal	Mean Squared Error	-14.544	2.622e-47	-32.452	6.595e-217
Per Dataset	Mean Absolute Error	17.553	1.102e-67	2.333	1.965e-02
Per Dataset	Mean Squared Error	-2.630	8.547e-03	-33.449	1.928e-229

Table 13: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from PulseDB vs No pretrain) for SBP prediction with Alexnet.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-17.204	3.842e-65	-39.638	0.000
Per Signal	Mean Squared Error	-28.756	5.491e-173	-14.122	9.718e-45
Per Dataset	Mean Absolute Error	-23.561	1.169e-118	4.182	2.925e-05
Per Dataset	Mean Squared Error	7.355	2.100e-13	-15.165	3.204e-51

Table 14: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from PulseDB vs No pretrain) for DBP prediction with Alexnet.

ResNet

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-10.120	6.291e-24	14.954	7.118e-50
Per Signal	Mean Squared Error	31.958	8.501e-211	-26.544	9.512e-149
Per Dataset	Mean Absolute Error	-17.886	3.663e-70	2.518	6.045e-03
Per Dataset	Mean Squared Error	-2.078	2.810e-02	-15.493	2.359e-53

Table 15: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from PulseDB vs No pretrain) for SBP prediction with ResNet.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	51.365	0.000	9.810	1.380e-22
Per Signal	Mean Squared Error	-21.517	7.859e-100	-18.547	3.431e-75
Per Dataset	Mean Absolute Error	-22.891	2.567e-112	-7.977	1.715e-15
Per Dataset	Mean Squared Error	32.179	1.639e-213	-56.326	0.000

Table 16: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from PulseDB vs No pretrain) for DBP prediction with ResNet.

BiLSTM

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-10.055	1.214e-23	-10.247	1.745e-24
Per Signal	Mean Squared Error	2.149	2.505e-02	13.633	7.569e-42
Per Dataset	Mean Absolute Error	43.615	0.000	29.880	6.745e-186
Per Dataset	Mean Squared Error	-45.648	0.000	-3.289	1.009e-03

Table 17: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from PulseDB vs No pretrain) for SBP prediction with BiLSTM.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-26.432	1.439e-147	-3.034	2.420e-03
Per Signal	Mean Squared Error	31.291	1.154e-202	-11.100	2.041e-28
Per Dataset	Mean Absolute Error	-3.136	1.722e-03	19.232	1.409e-80
Per Dataset	Mean Squared Error	-17.753	3.630e-69	6.940	4.238e-12

Table 18: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from PulseDB vs No pretrain) for DBP prediction with BiLSTM.

Slapničar et al.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-7.009	2.607e-12	39.818	0.000
Per Signal	Mean Squared Error	-4.690	2.781e-06	26.582	3.724e-149
Per Dataset	Mean Absolute Error	25.407	6.385e-137	-5.040	4.758e-07
Per Dataset	Mean Squared Error	12.626	3.402e-36	14.529	3.224e-47

Table 19: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from PulseDB vs No pretrain) for SBP prediction with Slapničar et al..

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	5.159	2.544e-07	58.179	0.000
Per Signal	Mean Squared Error	17.395	1.583e-66	21.276	1.039e-97
Per Dataset	Mean Absolute Error	13.372	2.419e-40	-8.046	9.840e-16
Per Dataset	Mean Squared Error	9.791	1.648e-22	13.187	2.713e-39

Table 20: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from PulseDB vs No pretrain) for DBP prediction with Slapničar et al..

BP-Net

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	6.369	2.012e-10	2.847	4.430e-03
Per Signal	Mean Squared Error	-7.568	4.224e-14	3.167	1.548e-03
Per Dataset	Mean Absolute Error	-4.095	4.256e-05	8.887	7.656e-19
Per Dataset	Mean Squared Error	16.992	1.285e-63	-2.550	1.079e-02

Table 21: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from PulseDB vs No pretrain) for SBP prediction with BP-Net.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	9.790	1.680e-22	-2.982	3.261e-02
Per Signal	Mean Squared Error	11.648	4.216e-31	9.460	3.990e-21
Per Dataset	Mean Absolute Error	2.060	3.944e-02	22.979	3.829e-113
Per Dataset	Mean Squared Error	7.718	1.328e-14	-2.515	6.065e-03

Table 22: Results of the paired t-tests based on pretrain setting (pretrained with MIMIC-III from PulseDB vs No pretrain) for DBP prediction with BP-Net.

3.2 Results of the paired t-tests based on HR feature

Alexnet

Pretrain Dataset	Normalization	Loss Function	$T_c = 1.96 / -1.96$	
			T-value	p-value
MIMIC-III from Kaggle	Per Signal	Mean Absolute Error	8.120	5.369e-16
MIMIC-III from Kaggle	Per Signal	Mean Squared Error	9.694	4.248e-22
MIMIC-III from Kaggle	Per Dataset	Mean Absolute Error	27.151	3.102e-155
MIMIC-III from Kaggle	Per Dataset	Mean Squared Error	-7.811	6.393e-15
MIMIC-III from PulseDB	Per Signal	Mean Absolute Error	8.754	2.489e-18
MIMIC-III from PulseDB	Per Signal	Mean Squared Error	11.678	2.996e-31
MIMIC-III from PulseDB	Per Dataset	Mean Absolute Error	-7.171	8.127e-13
MIMIC-III from PulseDB	Per Dataset	Mean Squared Error	-23.447	1.438e-117

Table 23: Results of the paired t-tests based on HR feature for SBP prediction with Alexnet.

Pretrain Dataset	Normalization	Loss Function	$T_c = 1.96 / -1.96$	
			T-value	p-value
MIMIC-III from Kaggle	Per Signal	Mean Absolute Error	-8.381	6.164e-17
MIMIC-III from Kaggle	Per Signal	Mean Squared Error	0.940	3.472e-01
MIMIC-III from Kaggle	Per Dataset	Mean Absolute Error	6.624	3.739e-11
MIMIC-III from Kaggle	Per Dataset	Mean Squared Error	-33.378	1.528e-228
MIMIC-III from PulseDB	Per Signal	Mean Absolute Error	-4.561	5.159e-06
MIMIC-III from PulseDB	Per Signal	Mean Squared Error	14.064	2.167e-44
MIMIC-III from PulseDB	Per Dataset	Mean Absolute Error	0.271	7.861e-01
MIMIC-III from PulseDB	Per Dataset	Mean Squared Error	-37.746	1.152e-286

Table 24: Results of the paired t-tests based on HR feature for DBP prediction with Alexnet.

ResNet

Pretrain Dataset	Normalization	Loss Function	$T_c = 1.96 / -1.96$	
			T-value	p-value
MIMIC-III from Kaggle	Per Signal	Mean Absolute Error	21.341	2.830e-98
MIMIC-III from Kaggle	Per Signal	Mean Squared Error	3.442	5.803e-04
MIMIC-III from Kaggle	Per Dataset	Mean Absolute Error	15.637	2.661e-54
MIMIC-III from Kaggle	Per Dataset	Mean Squared Error	9.509	2.502e-21
MIMIC-III from PulseDB	Per Signal	Mean Absolute Error	9.421	5.790e-21
MIMIC-III from PulseDB	Per Signal	Mean Squared Error	-3.118	1.827e-03
MIMIC-III from PulseDB	Per Dataset	Mean Absolute Error	25.846	1.964e-141
MIMIC-III from PulseDB	Per Dataset	Mean Squared Error	-4.404	1.076e-05

Table 25: Results of the paired t-tests based on HR feature for SBP prediction with ResNet.

Pretrain Dataset	Normalization	Loss Function	$T_c = 1.96 / -1.96$	
			T-value	p-value
MIMIC-III from Kaggle	Per Signal	Mean Absolute Error	25.618	4.448e-139
MIMIC-III from Kaggle	Per Signal	Mean Squared Error	6.118	9.911e-10
MIMIC-III from Kaggle	Per Dataset	Mean Absolute Error	3.801	1.452e-04
MIMIC-III from Kaggle	Per Dataset	Mean Squared Error	-15.465	3.602e-53
MIMIC-III from PulseDB	Per Signal	Mean Absolute Error	-9.842	1.002e-22
MIMIC-III from PulseDB	Per Signal	Mean Squared Error	17.313	6.239e-66
MIMIC-III from PulseDB	Per Dataset	Mean Absolute Error	6.247	4.410e-10
MIMIC-III from PulseDB	Per Dataset	Mean Squared Error	-28.805	1.536e-173

Table 26: Results of the paired t-tests based on HR feature for DBP prediction with ResNet.

BiLSTM

Pretrain Dataset	Normalization	Loss Function	$T_c = 1.96 / -1.96$	
			T-value	p-value
MIMIC-III from Kaggle	Per Signal	Mean Absolute Error	1.669	9.522e-02
MIMIC-III from Kaggle	Per Signal	Mean Squared Error	-5.830	5.753e-09
MIMIC-III from Kaggle	Per Dataset	Mean Absolute Error	19.777	5.329e-85
MIMIC-III from Kaggle	Per Dataset	Mean Squared Error	4.078	4.590e-05
MIMIC-III from PulseDB	Per Signal	Mean Absolute Error	10.677	1.982e-26
MIMIC-III from PulseDB	Per Signal	Mean Squared Error	2.501	1.240e-02
MIMIC-III from PulseDB	Per Dataset	Mean Absolute Error	15.116	6.552e-51
MIMIC-III from PulseDB	Per Dataset	Mean Squared Error	9.021	2.306e-19

Table 27: Results of the paired t-tests based on HR feature for SBP prediction with BiLSTM.

Pretrain Dataset	Normalization	Loss Function	$T_c = 1.96 / -1.96$	
			T-value	p-value
MIMIC-III from Kaggle	Per Signal	Mean Absolute Error	15.973	1.524e-56
MIMIC-III from Kaggle	Per Signal	Mean Squared Error	-0.651	5.152e-01
MIMIC-III from Kaggle	Per Dataset	Mean Absolute Error	26.496	3.089e-148
MIMIC-III from Kaggle	Per Dataset	Mean Squared Error	-17.328	4.872e-66
MIMIC-III from PulseDB	Per Signal	Mean Absolute Error	24.218	5.122e-125
MIMIC-III from PulseDB	Per Signal	Mean Squared Error	-19.221	1.719e-80
MIMIC-III from PulseDB	Per Dataset	Mean Absolute Error	24.411	6.372e-127
MIMIC-III from PulseDB	Per Dataset	Mean Squared Error	-7.625	2.734e-14

Table 28: Results of the paired t-tests based on HR feature for DBP prediction with BiLSTM.

Slapničar et al.

Pretrain Dataset	Normalization	Loss Function	$T_c = 1.96 / -1.96$	
			T-value	p-value
MIMIC-III from Kaggle	Per Signal	Mean Absolute Error	21.098	3.699e-96
MIMIC-III from Kaggle	Per Signal	Mean Squared Error	19.436	3.210e-82
MIMIC-III from Kaggle	Per Dataset	Mean Absolute Error	-11.584	8.819e-31
MIMIC-III from Kaggle	Per Dataset	Mean Squared Error	-31.377	1.058e-203
MIMIC-III from PulseDB	Per Signal	Mean Absolute Error	35.053	3.215e-250
MIMIC-III from PulseDB	Per Signal	Mean Squared Error	36.358	1.227e-267
MIMIC-III from PulseDB	Per Dataset	Mean Absolute Error	-22.628	7.154e-110
MIMIC-III from PulseDB	Per Dataset	Mean Squared Error	16.289	1.074e-58

Table 29: Results of the paired t-tests based on HR feature for SBP prediction with Slapničar et al..

Pretrain Dataset	Normalization	Loss Function	$T_c = 1.96 / -1.96$	
			T-value	p-value
MIMIC-III from Kaggle	Per Signal	Mean Absolute Error	17.035	6.343e-64
MIMIC-III from Kaggle	Per Signal	Mean Squared Error	-8.463	3.083e-17
MIMIC-III from Kaggle	Per Dataset	Mean Absolute Error	12.022	5.277e-33
MIMIC-III from Kaggle	Per Dataset	Mean Squared Error	-14.889	1.818e-49
MIMIC-III from PulseDB	Per Signal	Mean Absolute Error	-7.415	1.342e-13
MIMIC-III from PulseDB	Per Signal	Mean Squared Error	32.664	1.526e-219
MIMIC-III from PulseDB	Per Dataset	Mean Absolute Error	-14.267	1.308e-45
MIMIC-III from PulseDB	Per Dataset	Mean Squared Error	15.445	4.866e-53

Table 30: Results of the paired t-tests based on HR feature for DBP prediction with Slapničar et al..

BP-Net

Pretrain Dataset	Normalization	Loss Function	$T_c = 1.96 / -1.96$	
			T-value	p-value
MIMIC-III from Kaggle	Per Signal	Mean Absolute Error	0.354	7.230e-01
MIMIC-III from Kaggle	Per Signal	Mean Squared Error	2.086	3.703e-02
MIMIC-III from Kaggle	Per Dataset	Mean Absolute Error	-1.424	1.544e-01
MIMIC-III from Kaggle	Per Dataset	Mean Squared Error	-1.081	2.797e-01
MIMIC-III from PulseDB	Per Signal	Mean Absolute Error	2.511	1.206e-02
MIMIC-III from PulseDB	Per Signal	Mean Squared Error	2.955	3.139e-03
MIMIC-III from PulseDB	Per Dataset	Mean Absolute Error	1.818	6.916e-02
MIMIC-III from PulseDB	Per Dataset	Mean Squared Error	-11.182	8.228e-29

Table 31: Results of the paired t-tests based on HR feature for SBP prediction with BP-Net.

Pretrain Dataset	Normalization	Loss Function	$T_c = 1.96 / -1.96$	
			T-value	p-value
MIMIC-III from Kaggle	Per Signal	Mean Absolute Error	-12.190	7.096e-34
MIMIC-III from Kaggle	Per Signal	Mean Squared Error	-8.100	6.347e-16
MIMIC-III from Kaggle	Per Dataset	Mean Absolute Error	9.055	1.704e-19
MIMIC-III from Kaggle	Per Dataset	Mean Squared Error	-0.583	5.601e-01
MIMIC-III from PulseDB	Per Signal	Mean Absolute Error	-6.743	1.663e-11
MIMIC-III from PulseDB	Per Signal	Mean Squared Error	-8.130	4.963e-16
MIMIC-III from PulseDB	Per Dataset	Mean Absolute Error	8.033	1.086e-15
MIMIC-III from PulseDB	Per Dataset	Mean Squared Error	-3.629	2.864e-04

Table 32: Results of the paired t-tests based on HR feature for DBP prediction with BP-Net.

3.3 Results of the paired t-tests based on pretrain dataset (pretrained on MIMIC-III from Kaggle vs pretrained on MIMIC-III from PulseDB)

Alexnet

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	4.968	6.909e-07	5.033	4.947e-07
Per Signal	Mean Squared Error	-18.029	3.108e-71	-14.831	4.221e-49
Per Dataset	Mean Absolute Error	8.851	1.062e-18	-14.937	9.079e-50
Per Dataset	Mean Squared Error	16.504	3.533e-60	-2.612	9.030e-03

Table 33: Results of the paired t-tests based on pretrain dataset for SBP prediction with Alexnet.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-8.001	1.407e-15	-3.322	8.964e-04
Per Signal	Mean Squared Error	-38.490	4.641e-297	-22.561	2.916e-109
Per Dataset	Mean Absolute Error	8.518	1.926e-17	2.132	3.305e-02
Per Dataset	Mean Squared Error	22.629	6.881e-110	30.699	1.490e-195

Table 34: Results of the paired t-tests based on pretrain dataset for DBP prediction with Alexnet.

ResNet

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	13.840	4.646e-43	-6.419	1.448e-10
Per Signal	Mean Squared Error	12.566	7.166e-36	4.298	1.741e-05
Per Dataset	Mean Absolute Error	14.911	1.339e-49	24.879	1.423e-131
Per Dataset	Mean Squared Error	-0.084	9.334e-01	-11.733	1.575e-31

Table 35: Results of the paired t-tests based on pretrain dataset for SBP prediction with ResNet.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	31.084	3.687e-200	-4.230	2.360e-05
Per Signal	Mean Squared Error	20.363	7.264e-90	29.656	2.755e-183
Per Dataset	Mean Absolute Error	-22.332	3.754e-107	-15.962	1.813e-56
Per Dataset	Mean Squared Error	1.162	2.454e-01	-24.910	6.958e-132

Table 36: Results of the paired t-tests based on pretrain dataset for DBP prediction with ResNet.

BiLSTM

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-8.161	3.850e-16	5.513	3.630e-08
Per Signal	Mean Squared Error	-5.063	4.231e-07	6.825	9.464e-12
Per Dataset	Mean Absolute Error	20.181	2.435e-88	8.138	4.634e-16
Per Dataset	Mean Squared Error	-11.575	9.837e-31	-1.690	9.107e-02

Table 37: Results of the paired t-tests based on pretrain dataset for SBP prediction with BiLSTM.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-18.507	6.953e-75	-1.138	2.553e-01
Per Signal	Mean Squared Error	22.940	8.867e-113	-0.964	3.348e-01
Per Dataset	Mean Absolute Error	22.107	4.120e-105	16.728	9.571e-62
Per Dataset	Mean Squared Error	-16.378	2.617e-59	-5.821	6.070e-09

Table 38: Results of the paired t-tests based on pretrain dataset for DBP prediction with BiLSTM.

Slapničar et al.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	3.829	1.297e-04	25.564	1.582e-138
Per Signal	Mean Squared Error	-7.310	2.941e-13	6.953	3.863e-12
Per Dataset	Mean Absolute Error	22.841	7.520e-112	18.797	3.877e-77
Per Dataset	Mean Squared Error	-35.725	3.936e-259	13.957	9.493e-44

Table 39: Results of the paired t-tests based on pretrain dataset for SBP prediction with Slapničar et al..

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	48.936	0.000	26.238	1.600e-145
Per Signal	Mean Squared Error	-28.393	6.802e-169	15.298	4.410e-52
Per Dataset	Mean Absolute Error	15.087	1.013e-50	-21.991	4.637e-104
Per Dataset	Mean Squared Error	-27.625	2.143e-160	0.168	8.665e-01

Table 40: Results of the paired t-tests based on pretrain dataset for DBP prediction with Slapničar et al..

BP-Net

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	5.692	1.297e-08	7.994	1.493e-15
Per Signal	Mean Squared Error	-2.123	3.376e-02	-1.167	2.431e-01
Per Dataset	Mean Absolute Error	-2.374	1.763e-02	0.868	3.856e-01
Per Dataset	Mean Squared Error	11.558	1.194e-30	1.383	1.666e-01

Table 41: Results of the paired t-tests based on pretrain dataset for SBP prediction with BP-Net.

Normalization	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
Per Signal	Mean Absolute Error	-1.274	2.028e-01	3.763	1.693e-04
Per Signal	Mean Squared Error	4.693	2.742e-06	4.846	1.284e-06
Per Dataset	Mean Absolute Error	0.946	3.440e-01	-0.449	6.532e-01
Per Dataset	Mean Squared Error	0.844	3.987e-01	-2.203	2.761e-02

Table 42: Results of the paired t-tests based on pretrain dataset for DBP prediction with BP-Net.

3.4 Results of the paired t-tests based on normalization methods

Alexnet

Pretrain Dataset	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Mean Absolute Error	-0.063	9.500e-01	10.626	3.398e-26
MIMIC-III from Kaggle	Mean Squared Error	-3.235	1.221e-03	-17.533	1.531e-67
MIMIC-III from PulseDB	Mean Absolute Error	3.752	1.768e-04	-10.252	1.657e-24
MIMIC-III from PulseDB	Mean Squared Error	27.925	1.072e-163	-6.251	4.283e-10

Table 43: Results of the paired t-tests based on normalization methods for SBP prediction with Alexnet.

Pretrain Dataset	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Mean Absolute Error	-22.031	2.009e-104	-7.194	6.870e-13
MIMIC-III from Kaggle	Mean Squared Error	-31.531	1.425e-205	-44.377	0.000
MIMIC-III from PulseDB	Mean Absolute Error	-5.854	4.997e-09	-2.470	1.353e-02
MIMIC-III from PulseDB	Mean Squared Error	22.158	1.426e-105	-22.451	3.042e-108

Table 44: Results of the paired t-tests based on normalization methods for DBP prediction with Alexnet.

ResNet

Pretrain Dataset	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Mean Absolute Error	2.420	1.554e-02	-6.106	1.067e-09
MIMIC-III from Kaggle	Mean Squared Error	10.511	1.142e-25	13.428	1.161e-40
MIMIC-III from PulseDB	Mean Absolute Error	5.622	1.952e-08	25.992	5.919e-143
MIMIC-III from PulseDB	Mean Squared Error	-1.928	5.385e-02	-3.968	7.302e-05

Table 45: Results of the paired t-tests based on normalization methods for SBP prediction with ResNet.

Pretrain Dataset	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Mean Absolute Error	15.172	2.879e-51	-1.597	1.103e-01
MIMIC-III from Kaggle	Mean Squared Error	32.826	1.425e-221	13.878	2.778e-43
MIMIC-III from PulseDB	Mean Absolute Error	-28.663	6.255e-172	-12.541	9.811e-36
MIMIC-III from PulseDB	Mean Squared Error	10.377	4.572e-25	-31.689	1.687e-207

Table 46: Results of the paired t-tests based on normalization methods for DBP prediction with ResNet.

BiLSTM

Pretrain Dataset	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Mean Absolute Error	-20.681	1.461e-92	7.344	2.275e-13
MIMIC-III from Kaggle	Mean Squared Error	-14.264	1.362e-45	2.614	8.966e-03
MIMIC-III from PulseDB	Mean Absolute Error	0.324	7.460e-01	9.938	3.896e-23
MIMIC-III from PulseDB	Mean Squared Error	-17.725	5.846e-69	-5.901	3.753e-09

Table 47: Results of the paired t-tests based on normalization methods for SBP prediction with BiLSTM.

Pretrain Dataset	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Mean Absolute Error	-9.358	1.043e-20	8.867	9.156e-19
MIMIC-III from Kaggle	Mean Squared Error	14.297	8.605e-46	-11.753	1.252e-31
MIMIC-III from PulseDB	Mean Absolute Error	21.600	1.450e-100	28.555	1.042e-170
MIMIC-III from PulseDB	Mean Squared Error	-22.845	6.900e-112	-18.255	6.046e-73

Table 48: Results of the paired t-tests based on normalization methods for DBP prediction with BiLSTM.

Slapničar et al.

Pretrain Dataset	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Mean Absolute Error	5.859	4.857e-09	-31.272	1.987e-202
MIMIC-III from Kaggle	Mean Squared Error	29.976	5.179e-187	-21.397	9.110e-99
MIMIC-III from PulseDB	Mean Absolute Error	26.020	3.036e-143	-32.662	1.592e-219
MIMIC-III from PulseDB	Mean Squared Error	-1.560	1.189e-01	-15.557	9.039e-54

Table 49: Results of the paired t-tests based on normalization methods for SBP prediction with Slapničar et al..

Pretrain Dataset	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Mean Absolute Error	5.637	1.794e-08	2.280	2.262e-02
MIMIC-III from Kaggle	Mean Squared Error	5.396	7.025e-08	-4.194	2.773e-05
MIMIC-III from PulseDB	Mean Absolute Error	-13.284	7.668e-40	-31.430	2.403e-204
MIMIC-III from PulseDB	Mean Squared Error	-6.627	3.655e-11	-12.885	1.318e-37

Table 50: Results of the paired t-tests based on normalization methods for DBP prediction with Slapničar et al..

BP-Net

Pretrain Dataset	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Mean Absolute Error	6.079	1.268e-09	4.331	1.503e-05
MIMIC-III from Kaggle	Mean Squared Error	0.785	4.324e-01	-2.381	1.727e-02
MIMIC-III from PulseDB	Mean Absolute Error	-1.905	5.685e-02	-2.494	1.267e-02
MIMIC-III from PulseDB	Mean Squared Error	14.509	4.314e-47	0.225	8.223e-01

Table 51: Results of the paired t-tests based on normalization methods for SBP prediction with BP-Net.

Pretrain Dataset	Loss Function	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Mean Absolute Error	-7.366	1.930e-13	13.996	5.545e-44
MIMIC-III from Kaggle	Mean Squared Error	-4.144	3.447e-05	3.414	6.429e-04
MIMIC-III from PulseDB	Mean Absolute Error	-5.184	2.230e-07	9.625	8.295e-22
MIMIC-III from PulseDB	Mean Squared Error	-8.148	4.289e-16	-3.683	2.320e-04

Table 52: Results of the paired t-tests based on normalization methods for DBP prediction with BP-Net.

3.5 Results of the paired t-tests based on loss function

Alexnet

Pretrain Dataset	Normalization	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Per Signal	20.599	7.355e-92	18.651	5.386e-76
MIMIC-III from Kaggle	Per Dataset	17.319	5.602e-66	-14.901	1.527e-49
MIMIC-III from PulseDB	Per Signal	-6.295	3.237e-10	-2.864	4.191e-03
MIMIC-III from PulseDB	Per Dataset	18.628	8.044e-76	1.820	6.873e-02

Table 53: Results of the paired t-tests based on loss function for SBP prediction with Alexnet.

Pretrain Dataset	Normalization	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Per Signal	9.596	1.097e-21	20.317	1.770e-89
MIMIC-III from Kaggle	Per Dataset	-1.667	9.546e-02	-35.241	1.073e-252
MIMIC-III from PulseDB	Per Signal	-21.208	4.084e-97	0.328	7.428e-01
MIMIC-III from PulseDB	Per Dataset	12.174	8.571e-34	-21.884	4.212e-103

Table 54: Results of the paired t-tests based on loss function for DBP prediction with Alexnet.

ResNet

Pretrain Dataset	Normalization	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Per Signal	29.545	5.241e-182	-31.238	5.164e-202
MIMIC-III from Kaggle	Per Dataset	20.508	4.319e-91	13.568	1.818e-41
MIMIC-III from PulseDB	Per Signal	18.098	9.445e-72	-3.950	7.879e-05
MIMIC-III from PulseDB	Per Dataset	1.197	2.315e-01	-50.982	0.000

Table 55: Results of the paired t-tests based on loss function for SBP prediction with ResNet.

Pretrain Dataset	Normalization	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Per Signal	-141.429	0.000	-147.114	0.000
MIMIC-III from Kaggle	Per Dataset	-66.691	0.000	-92.487	0.000
MIMIC-III from PulseDB	Per Signal	-158.222	0.000	-61.698	0.000
MIMIC-III from PulseDB	Per Dataset	-26.990	1.677e-153	-77.237	0.000

Table 56: Results of the paired t-tests based on loss function for DBP prediction with ResNet.

BiLSTM

Pretrain Dataset	Normalization	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Per Signal	-4.421	9.940e-06	-16.656	3.076e-61
MIMIC-III from Kaggle	Per Dataset	11.239	4.355e-29	-37.527	1.236e-283
MIMIC-III from PulseDB	Per Signal	0.013	9.899e-01	-35.006	1.336e-249
MIMIC-III from PulseDB	Per Dataset	-21.061	7.867e-96	-40.258	0.000

Table 57: Results of the paired t-tests based on loss function for SBP prediction with BiLSTM.

Pretrain Dataset	Normalization	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Per Signal	-14.225	2.347e-45	-71.815	0.000
MIMIC-III from Kaggle	Per Dataset	10.204	2.710e-24	-91.967	0.000
MIMIC-III from PulseDB	Per Signal	20.165	3.333e-88	-81.196	0.000
MIMIC-III from PulseDB	Per Dataset	-39.367	1.685e-309	-147.937	0.000

Table 58: Results of the paired t-tests based on loss function for DBP prediction with BiLSTM.

Slapničar et al.

Pretrain Dataset	Normalization	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Per Signal	-5.223	1.805e-07	-13.687	3.697e-42
MIMIC-III from Kaggle	Per Dataset	19.555	3.470e-83	4.793	1.673e-06
MIMIC-III from PulseDB	Per Signal	-22.060	1.116e-104	-35.769	1.026e-259
MIMIC-III from PulseDB	Per Dataset	-37.020	1.173e-276	-0.741	4.589e-01

Table 59: Results of the paired t-tests based on loss function for SBP prediction with Slapničar et al..

Pretrain Dataset	Normalization	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Per Signal	-1.480	1.389e-01	-54.074	0.000
MIMIC-III from Kaggle	Per Dataset	-2.833	4.627e-03	-67.087	0.000
MIMIC-III from PulseDB	Per Signal	-91.429	0.000	-68.986	0.000
MIMIC-III from PulseDB	Per Dataset	-38.240	1.508e-293	-16.581	1.017e-60

Table 60: Results of the paired t-tests based on loss function for DBP prediction with Slapničar et al..

BP-Net

Pretrain Dataset	Normalization	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Per Signal	0.109	9.130e-01	1.855	6.357e-02
MIMIC-III from Kaggle	Per Dataset	-5.244	1.614e-07	-4.828	1.406e-06
MIMIC-III from PulseDB	Per Signal	-7.914	2.823e-15	-7.242	4.827e-13
MIMIC-III from PulseDB	Per Dataset	8.540	1.600e-17	-4.344	1.417e-05

Table 61: Results of the paired t-tests based on loss function for SBP prediction with BP-Net.

Pretrain Dataset	Normalization	with HR		without HR	
		$T_c = 1.96 / -1.96$ T-value	p-value	$T_c = 1.96 / -1.96$ T-value	p-value
MIMIC-III from Kaggle	Per Signal	-3.602	3.177e-04	0.482	6.298e-01
MIMIC-III from Kaggle	Per Dataset	-0.236	8.138e-01	-10.130	5.721e-24
MIMIC-III from PulseDB	Per Signal	2.440	1.469e-02	1.399	1.618e-01
MIMIC-III from PulseDB	Per Dataset	-0.372	7.099e-01	-12.096	2.207e-33

Table 62: Results of the paired t-tests based on loss function for DBP prediction with BP-Net.

3.6 Results of the ANOVA test based on model architecture

Pretrain Dataset	Normalization	Loss Function	with HR $F_c = 2.37$		without HR $F_c = 2.37$	
			F-value	p-value	F-value	p-value
MIMIC-III from Kaggle	Per Signal	Mean Absolute Error	$F(4,39250) = 141.077$	$6.034e-120$	$F(4,39250) = 135.641$	$2.626e-115$
MIMIC-III from Kaggle	Per Signal	Mean Squared Error	$F(4,39250) = 56.479$	$1.374e-47$	$F(4,39250) = 8.627$	$5.896e-07$
MIMIC-III from Kaggle	Per Dataset	Mean Absolute Error	$F(4,39250) = 152.609$	$8.806e-130$	$F(4,39250) = 152.385$	$1.367e-129$
MIMIC-III from Kaggle	Per Dataset	Mean Squared Error	$F(4,39250) = 133.960$	$7.162e-114$	$F(4,39250) = 139.497$	$1.345e-118$
MIMIC-III from PulseDB	Per Signal	Mean Absolute Error	$F(4,39250) = 122.177$	$8.283e-104$	$F(4,39250) = 152.385$	$1.367e-129$
MIMIC-III from PulseDB	Per Signal	Mean Squared Error	$F(4,39250) = 196.190$	$7.233e-167$	$F(4,39250) = 88.388$	$6.559e-75$
MIMIC-III from PulseDB	Per Dataset	Mean Absolute Error	$F(4,39250) = 126.193$	$3.068e-107$	$F(4,39250) = 268.562$	$3.838e-228$
MIMIC-III from PulseDB	Per Dataset	Mean Squared Error	$F(4,39250) = 136.461$	$5.244e-116$	$F(4,39250) = 102.881$	$2.600e-87$

Table 63: Effects of different models (5 models) on SBP prediction errors as functions of pretrain dataset (Kaggle / PulseDB); Normalization methods (per signal / per dataset); Loss function (MAE / MSE); With / without HR during training.

Pretrain Dataset	Normalization	Loss Function	with HR $F_c = 2.37$		without HR $F_c = 2.37$	
			F-value	p-value	F-value	p-value
MIMIC-III from Kaggle	Per Signal	Mean Absolute Error	$F(4,39250) = 85.956$	$7.915e-73$	$F(4,39250) = 207.187$	$3.334e-176$
MIMIC-III from Kaggle	Per Signal	Mean Squared Error	$F(4,39250) = 23.193$	$3.578e-19$	$F(4,39250) = 22.354$	$1.841e-18$
MIMIC-III from Kaggle	Per Dataset	Mean Absolute Error	$F(4,39250) = 227.156$	$3.911e-193$	$F(4,39250) = 213.919$	$6.487e-182$
MIMIC-III from Kaggle	Per Dataset	Mean Squared Error	$F(4,39250) = 81.211$	$9.145e-69$	$F(4,39250) = 896.255$	0.000
MIMIC-III from PulseDB	Per Signal	Mean Absolute Error	$F(4,39250) = 207.461$	$1.949e-176$	$F(4,39250) = 176.984$	$1.533e-150$
MIMIC-III from PulseDB	Per Signal	Mean Squared Error	$F(4,39250) = 160.353$	$2.216e-136$	$F(4,39250) = 28.687$	$7.657e-24$
MIMIC-III from PulseDB	Per Dataset	Mean Absolute Error	$F(4,39250) = 90.346$	$1.382e-76$	$F(4,39250) = 179.276$	$1.722e-152$
MIMIC-III from PulseDB	Per Dataset	Mean Squared Error	$F(4,39250) = 71.078$	$4.338e-60$	$F(4,39250) = 70.519$	$1.306e-59$

Table 64: Effects of different models (5 models) on DBP prediction errors as functions of pretrain dataset (Kaggle / PulseDB); Normalization methods (per signal / per dataset); Loss function (MAE / MSE); With / without HR during training.

While results of the ANOVA tests confirm that different model architectures can lead to significant difference in model performance, the large sample sizes typically used in model training can cause bias and amplify even small changes into significant results. Nonetheless, a large range of F-values (ranging from 8.627 to 896.255) is observed, suggesting future work with permutation correction can reduce the bias and uncover finer differentiation in the results.