Comparison of Nine Deep Regressors in Continuous Blood Pressure Estimation Using Single-Channel Photoplethysmograms under the PulseDB

2025.1.22 SII2025

Takumi Yamamoto[1,2], Suguru Kanoga[1], Yuta Sugiura[2]

[1]National Institute of Advanced Industrial Science and Technology (AIST), [2]Keio University

Blood Pressure (BP)

- Blood pressure (BP) :
  - Important parameter for the diagnosis of cardiovascular disease (CVDs)
    - CVDs: In 2020, it is estimated that 19.05 million deaths globally were attributed [1]
  - Continuous monitoring of BP is important

- Cuff-based device is usually used for measurement
  - Stressful [2]
  - Considerable amount of time for each measurement
  - Cannot measure with high temporal resolution

Cuff-based measurement device

[1] F. D. Fuchs and P. K. Whelton, “High blood pressure and cardiovascular disease,” Hypertension, vol. 75, no. 2, pp. 285–292, 2020.
[2] D. Levy, M. G. Larson, R. S. Vasan, W. B. Kannel, and K. K. L. Ho, “The progression from hypertension to congestive heart failure,” JAMA, vol. 275, no. 20, pp. 1557–1562, 199

BP Measurement: Cuff-less Approach

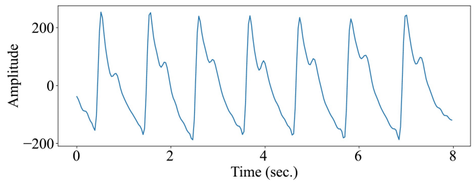
- Estimation from single-channel photoplethysmogram (PPG) sensors

- Deep learning-based models to estimate systolic and diastolic BPs (SBP, DBP)
  - SBP : Maximum BP, DBP: Minimum BP

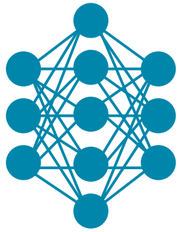
Cuff-less Approach

Wearable device with PPG sensor [3, 4]

PPG sensor data [5]



Deep learning model



Estimation

SBP (Maximum BP)
DBP (Minimum BP)

[3] Apple. Apple Watch. Apple Inc., Accessed 3 Jan. 2025, https://www.apple.com/jp/shop/buy-watch/apple-watch.
[4] Soxai. Soxai Ring 1. Soxai, Accessed 3 Jan. 2025, https://soxai.co.jp/products/soxai-ring-1.
[5] Kamei, Satoshi, et al. "One-Month Evaluation of Blood Pressure Estimation Using a Fine-Tuned Model with Wristband-Type Photoplethysmograms." IEEE Sensors Journal (2024).

Problem of Existing Research: Evaluation

- Comparison of proposed model to others is not fair across paper
  1. Dataset is different across papers
  2. Different methods of creating test dataset
     - Calibration-based and Calibration-free

Methods for creating test dataset

Calibration-based	Calibration-free
Same subject in training and test dataset	Different subject in training and test dataset

Contributions and Research Questions

Fair comparison of deep learning model using unified dataset under two different conditions (calibration-based and calibration-free conditions)

Research Questions

RQ1. **Is there a difference of BP estimation performance between calibration-based and calibration-free conditions?**

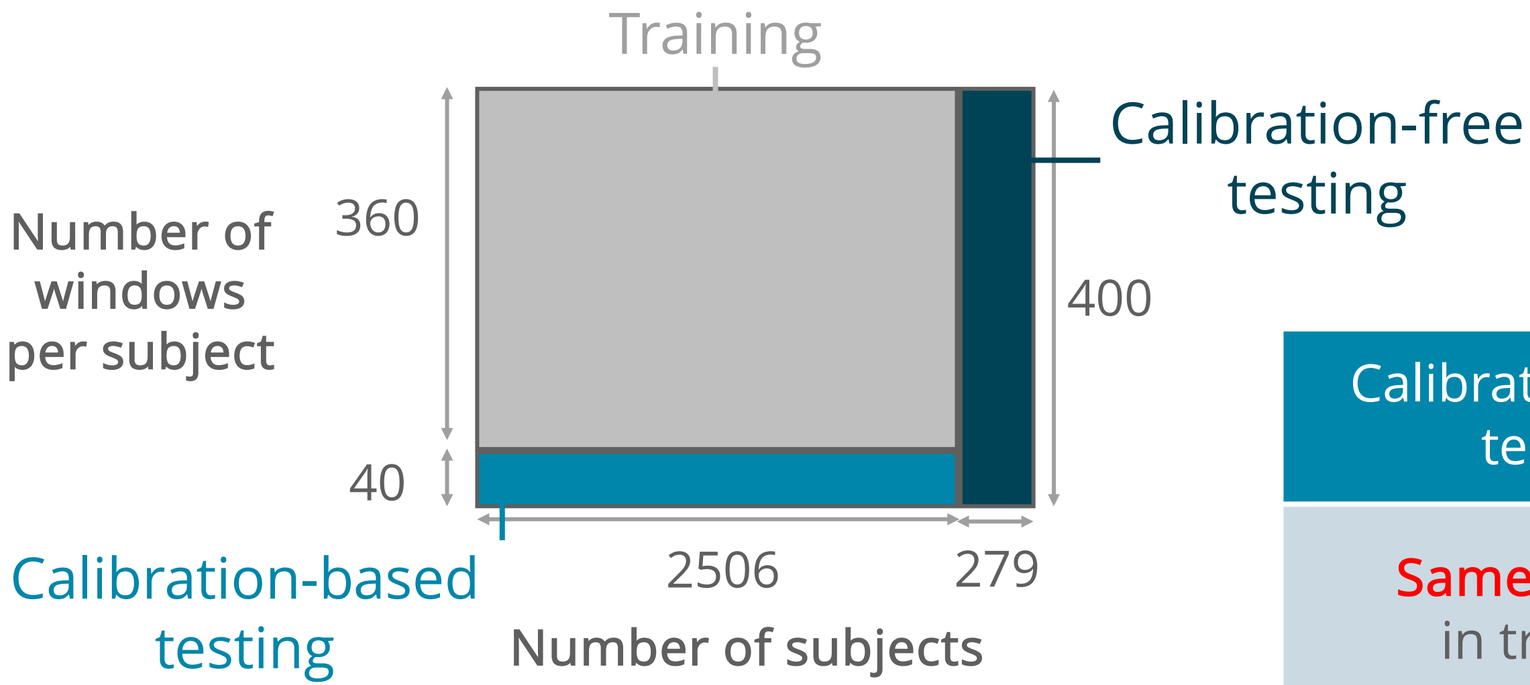
To provide guidance on evaluation methods for future researchers

RQ2. **Which deep learning models perform best under different conditions?**

To identify high-accuracy architectures using a unified dataset

Dataset: PulseDB[6]

- Contain data of 5361 subjects
- Already divided into training, calibration-based testing, calibration-free testing

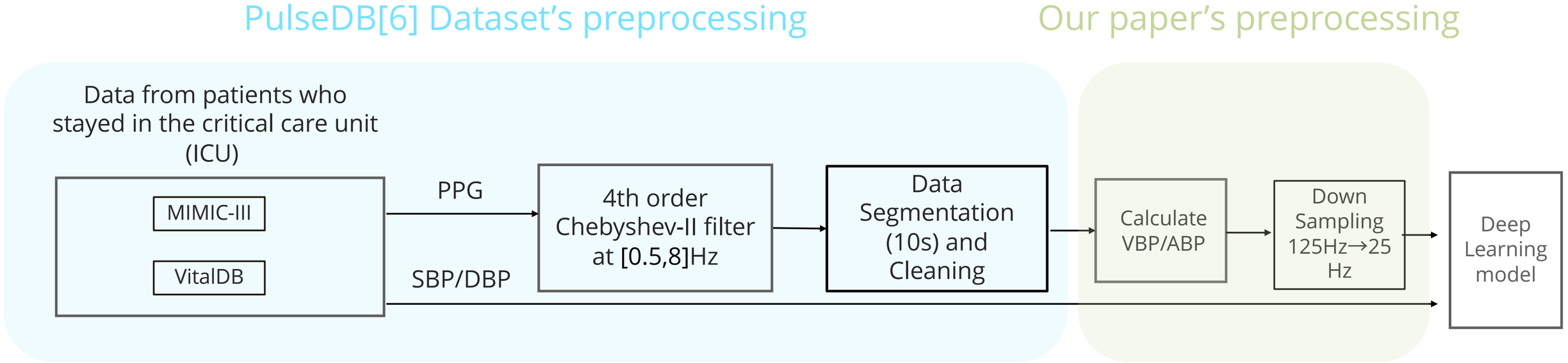


Calibration-based testing	Calibration-free testing
Same subject in training	Different subject in training and test dataset

[6] W. Wang, P. Mohseni, K. L. Kilgore, and L. Najafizadeh, “PulseDB: A large, cleaned dataset based on MIMIC-III and VitalDB for benchmarking cuff-less blood pressure estimation methods,” Front. Digit. Health, vol. 4, 2023.

Dataset: Data Preprocessing

- Pre-processing（data cleaning, filters) has already been carried out
- First derivative of PPG (velocity PPG: VPG) and second derivative of PPG (acceleration PPG: APG) were calculated
- Downsampling on 25 Hz to reduce computational cost

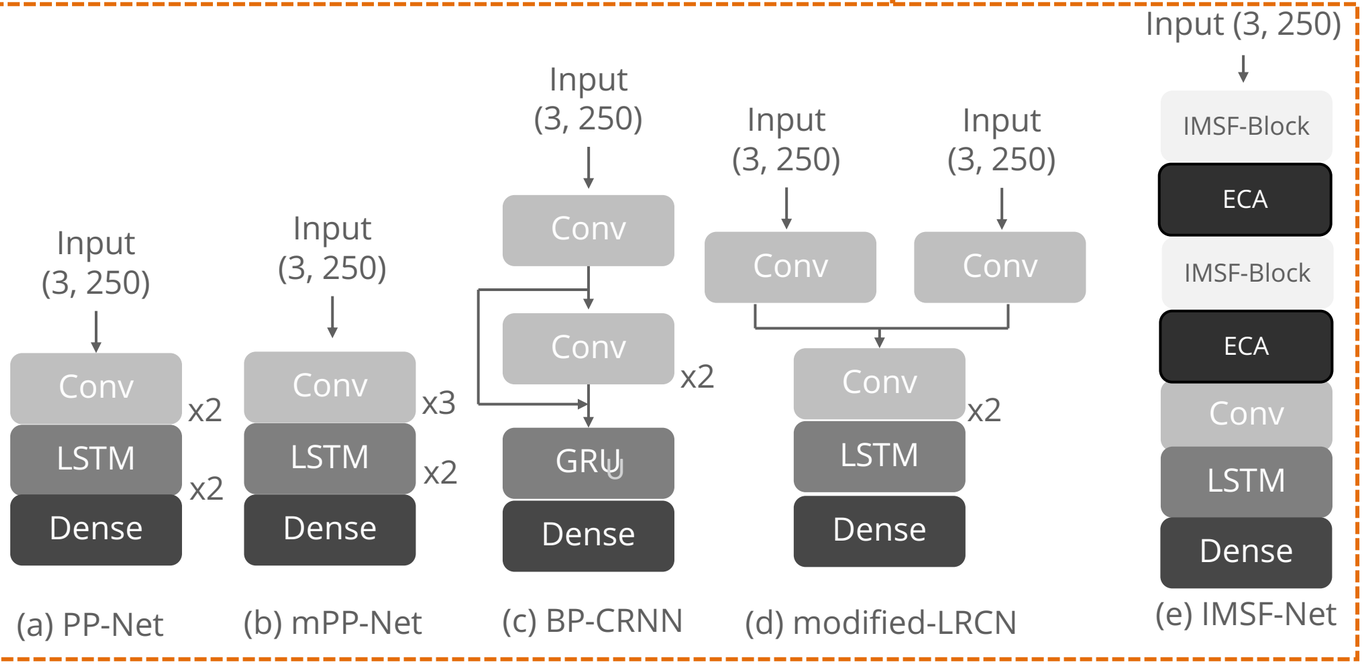


[6] W. Wang, P. Mohseni, K. L. Kilgore, and L. Najafizadeh, "PulseDB: A large, cleaned dataset based on MIMIC-III and VitalDB for benchmarking cuff-less blood pressure estimation methods," Front. Digit. Health, vol. 4, 2023.

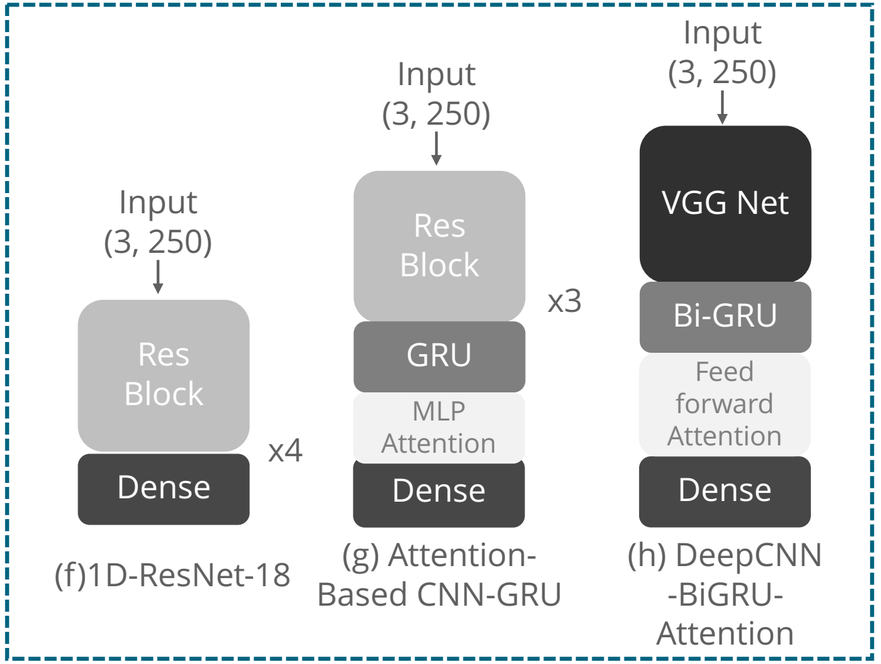
[7] Panwar, Madhuri, et al. "PP-Net: A deep learning framework for PPG-based blood pressure and heart rate estimation." *IEEE Sensors Journal* 20.17 (2020): 10000-10011.

Nine Model's Architecture

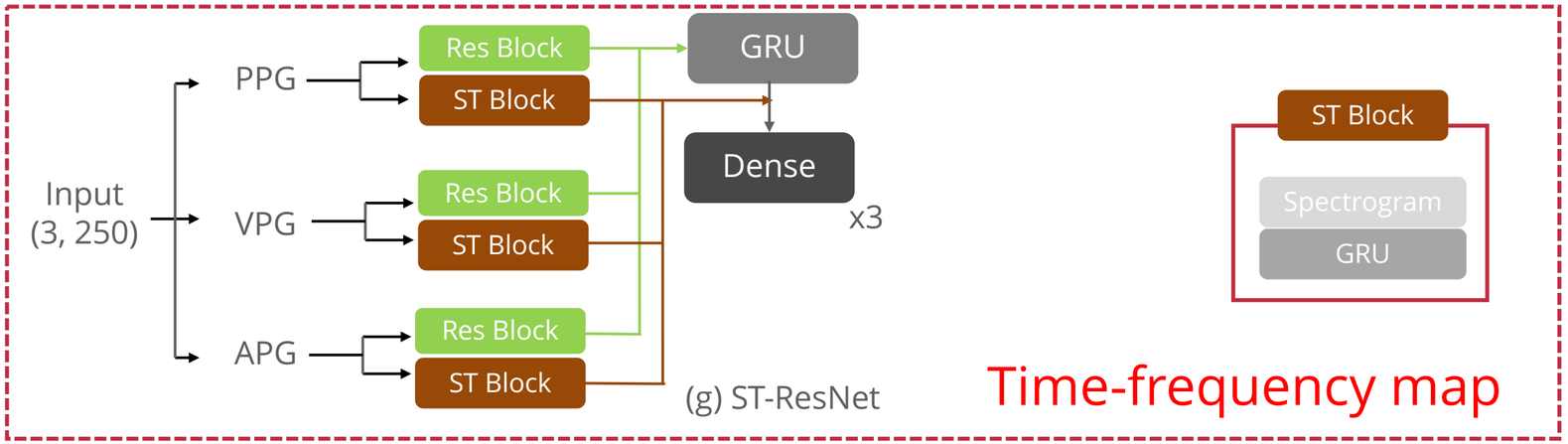
Shallow Conv Group



Deep Conv Group



Other



Conditions for Learning Parameters and Evaluation Metrics

- Two conditions for learning parameters
  - With weight decay (= 0.001), without weight decay (= 0)
  - Weight decay: regularization method for preventing the overfitting

- Evaluation metrics
  - Mean absolute error (MAE) between true and estimated SBP and DBP (mmHg)

Results: MAE (Mean Absolute Error)

- Calibration-based vs free: Approximately, calibration-based shows better performance than calibration-free
  - Calibration-based: Models without weight decay performed better
  - Calibration-free: Models with weight decay performed better, which **showed improved generalization**
- ST-ResNet approximately achieved the best performance

Model	Calibration-based				Calibration-free			
	w weight decay		w/o weight decay		w weight decay		w/o weight decay	
	SBP	DBP	SBP	DBP	SBP	DBP	SBP	DBP
PP-Net	15.01	9.00	13.80	8.28	14.75	8.94	14.09	8.56
mPP-Net	14.98	8.91	13.33	8.00	14.74	8.87	**14.01**	8.58
BP-CRNN	14.28	8.57	13.61	8.18	14.29	8.54	14.14	**8.55**
m-LRCN	14.70	8.84	13.53	8.20	14.44	8.84	14.74	9.14
IMSF-Net	13.61	8.16	12.38	7.39	13.85	8.45	15.46	9.55
1D ResNet18	13.59	8.18	12.77	7.72	13.94	8.45	15.98	9.85
Attention-Based CNN-GRU	14.53	8.68	12.50	7.49	14.37	8.70	15.31	9.51
DeepCNN-BiGRU-Attention	13.60	8.18	11.76	7.09	13.85	**8.43**	15.99	9.76
ST-ResNet	**13.52**	**8.09**	**10.67**	**6.37**	**13.84**	**8.43**	15.32	9.59

Shallow Conv Group

Deep Conv Group

Other Group

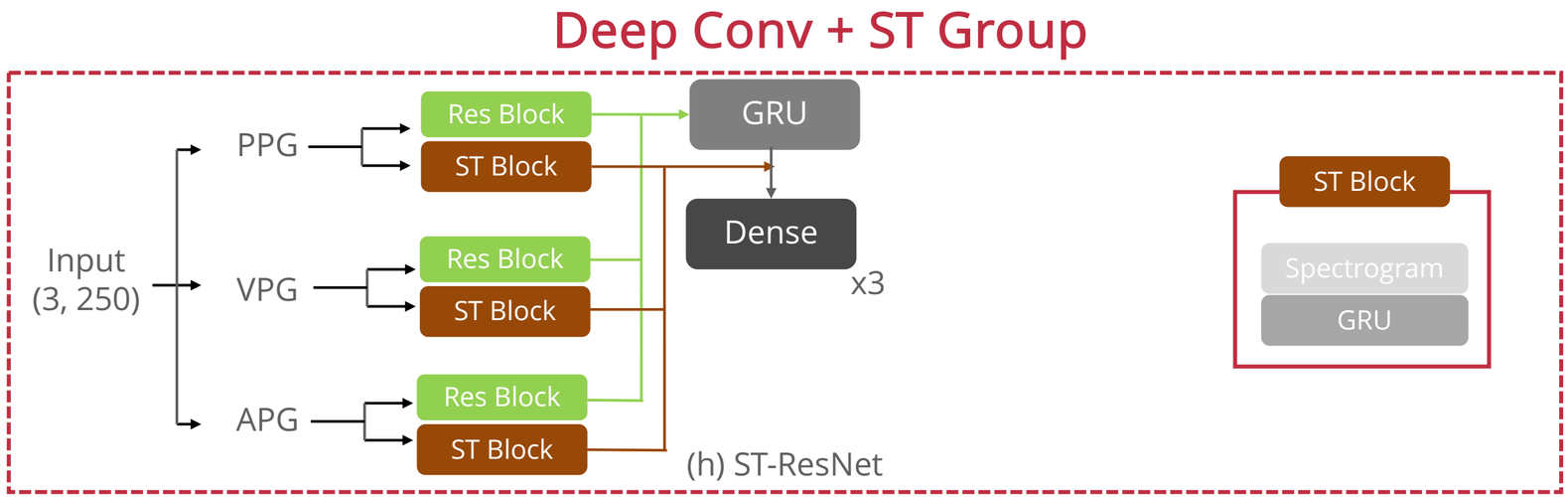
Responses to Research Questions

RQ1: Is there a difference between calibration-based and calibration-free performance?

- Yes, calibration-based subsets showed better performance than calibration-free subsets
- Weight decay improved calibration-free results by reducing overfitting risks

RQ2: Which model performed best?

- ST-ResNet performed the best across both subsets, especially with weight decay



Future Work and Takeaways

Future Directions:

- More complex architecture
- Explore transfer learning, continual learning

Takeaways for Future Researchers:

1. The model should be separately evaluated with calibration-based and calibration-free subset

2. To improve the accuracy for calibration-free subset, the model training should be prevented from the overfitting by using a weight decay, etc.

Summary

Comparison of Nine Deep Regressors in Continuous Blood Pressure Estimation Using Single-Channel Photoplethysmograms under the PulseDB

2025.1.22 SII2025

Takumi Yamamoto[1,2], Suguru Kanoga[1], Yuta Sugiura[2]
[1]National Institute of Advanced Industrial Science and Technology (AIST), [2]Keio University

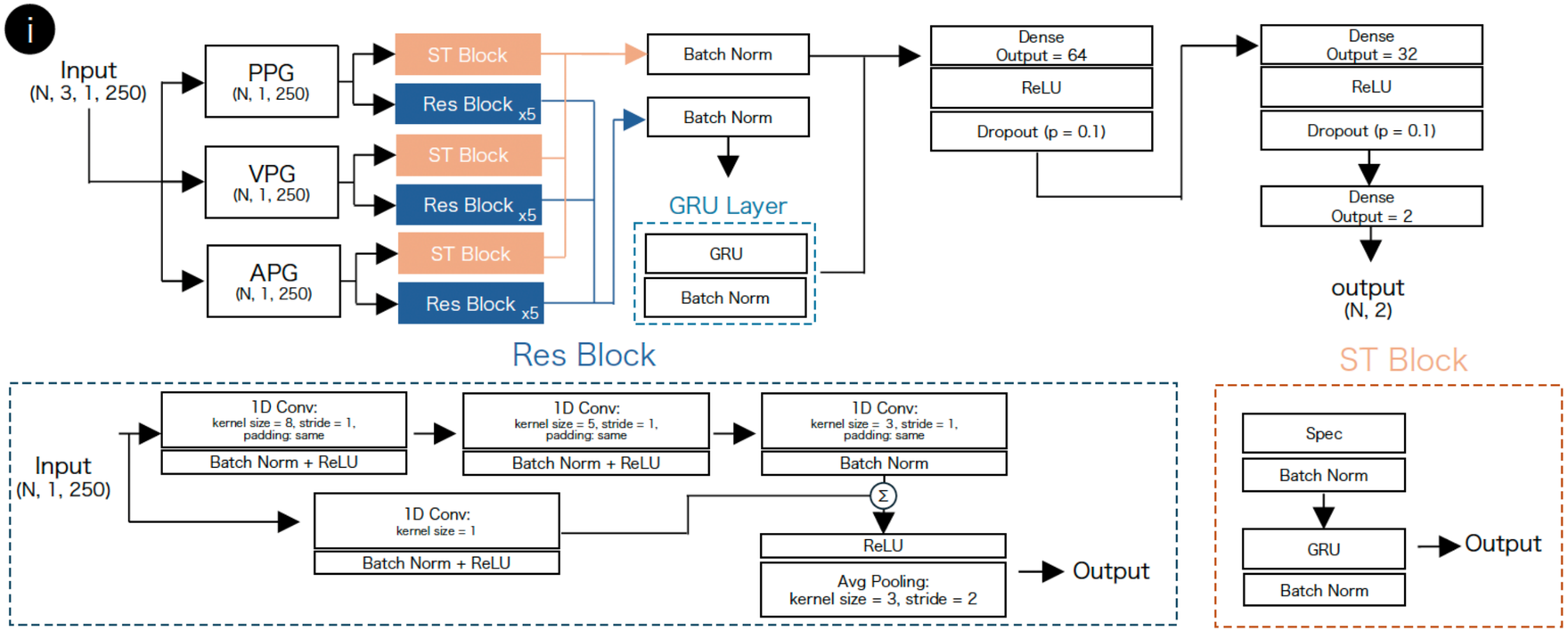
Background	PPG-based model to estimate the SBP and DBP
Problem	Comparison with different dataset Different methods of splitting training and testing subsets
Proposal	Fair comparison of deep learning model using unified dataset in calibration-based/free conditions
Results	Calibration-based showed better performance than calibration-free Weight decay improved calibration-free, ST-ResNet performed the best
Takeaways	1. The model should be separately evaluated with calibration-based and calibration-free subset 2. To improve the accuracy for calibration-free subset, the model training should be prevented from the overfitting by using a weight decay, etc.

This work was partially supported by JSPS KAKENHI Grant-in-Aid for Scientific Research~(B), Grant Number JP23H03445 and JP23K28135.

Supplemental Slides

Other: ST-ResNet (Spectral–temporal residual network) [7]

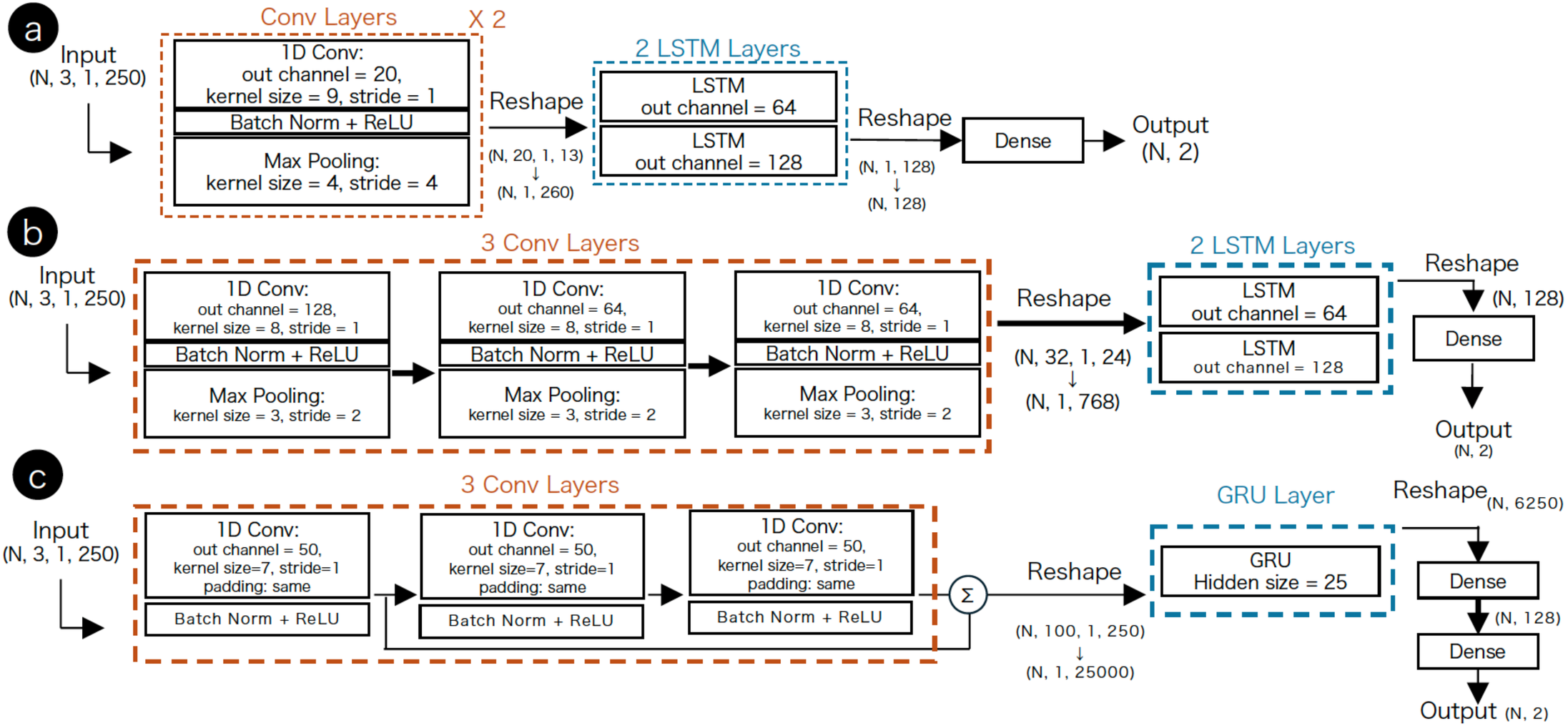
- Res Block: Using multiple convolutional layer with residual connection
- ST Block: Calculate spectrogram and GRU Layer



ST-ResNet

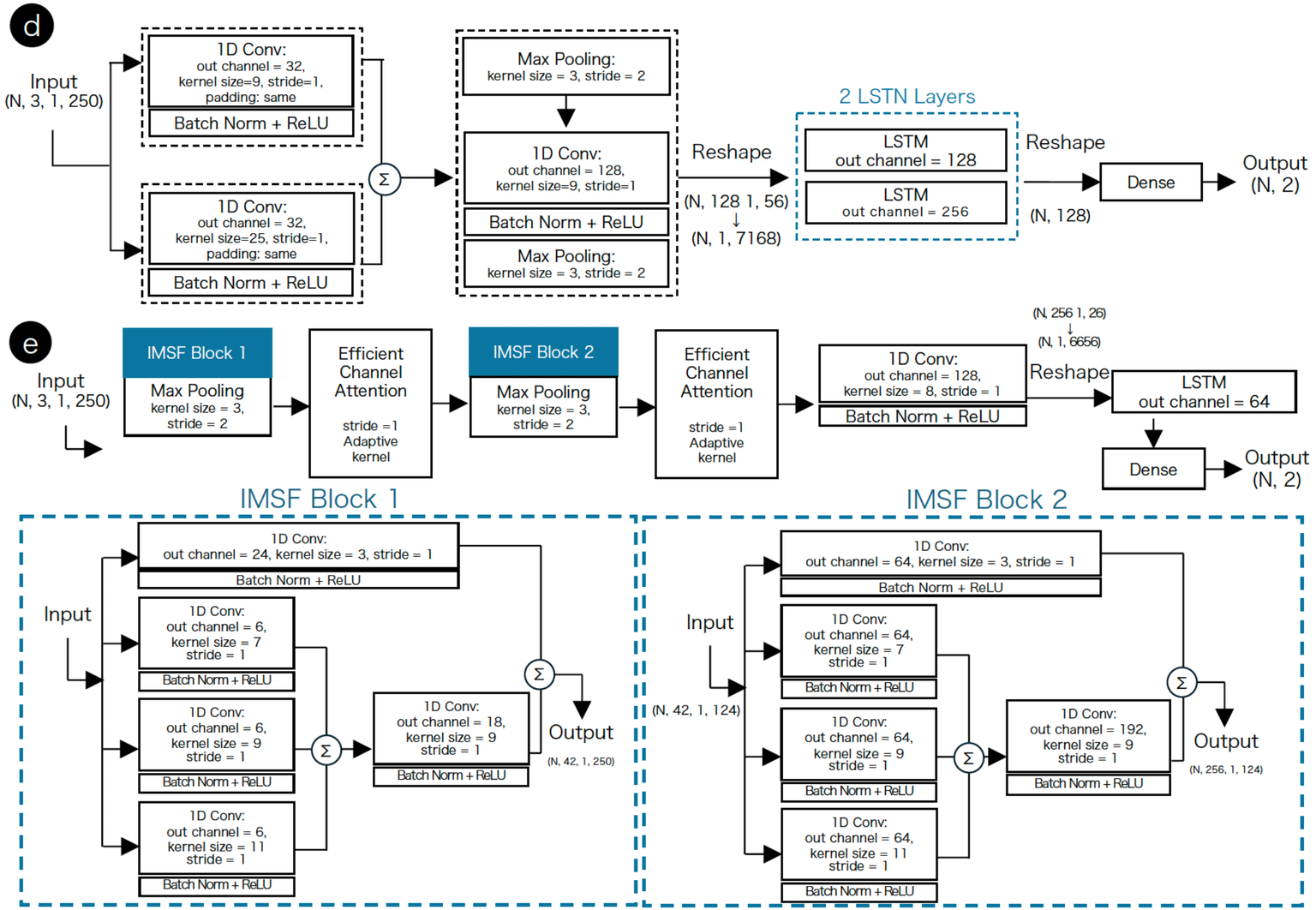
[7] Slapničar, Gašper, Nejc Mlakar, and Mitja Luštrek. "Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network." *Sensors* 19.15 (2019): 3420.

Shallow Conv Group (1/2)

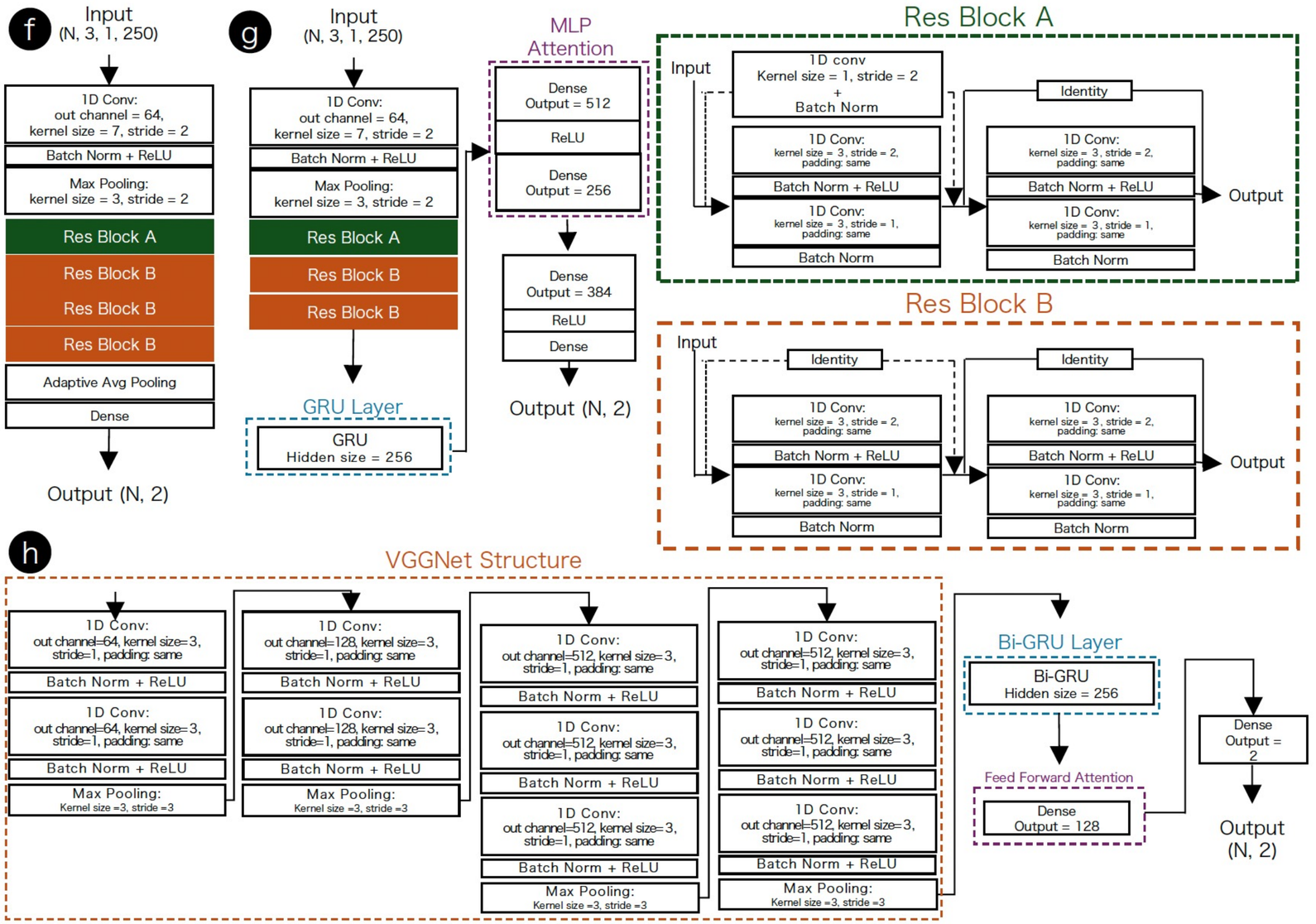


[7] Slapničar, Gašper, Nejc Mlakar, and Mitja Luštrek. "Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network." *Sensors* 19.15 (2019): 3420.

Shallow Conv Group (2/2)

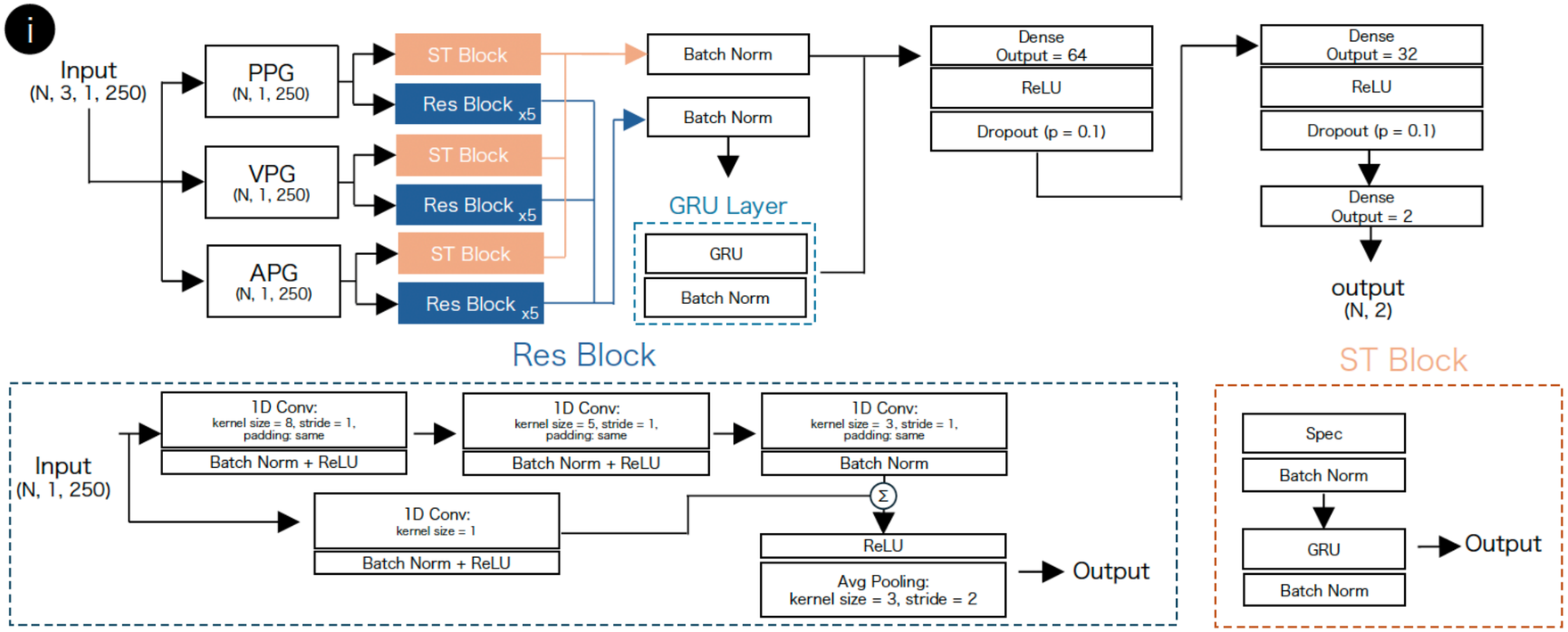


Deep Conv Group



Other: ST-ResNet (Spectral–temporal residual network) [7]

- Res Block: Using multiple convolutional layer with residual connection
- ST Block: Calculate spectrogram (time frequency map) and GRU Layer



ST-ResNet

[7] Slapničar, Gašper, Nejc Mlakar, and Mitja Luštrek. "Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network." *Sensors* 19.15 (2019): 3420.

Results: BHS Standard

- British Hypertension Society (BHS) Standard [8] :
  - Standard for BP measurement device: Grade A is necessary for medical devices
- Result
  - Calibration-based: ST-ResNet reached Grade B for DBP
  - Calibration-free: No models met Grade C for SBP and DBP

Calibration-based

Model	w weight decay								w/o weight decay							
	SBP				DBP				SBP				DBP			
	≦ 5	≦ 10	≦ 15	Grade	≦ 5	≦ 10	≦ 15	Grade	≦ 5	≦ 10	≦ 15	Grade	≦ 5	≦ 10	≦ 15	Grade
PP-Net	21.19	40.80	58.18	–	34.80	63.54	82.46	–	23.35	44.67	62.52	–	38.10	67.79	85.85	–
mPP-Net	21.23	40.83	58.38	–	35.13	64.03	82.88	–	24.22	46.30	64.27	–	39.42	69.52	87.01	–
BP-CRNN	22.55	43.47	60.78	–	36.48	65.86	84.60	–	23.83	45.55	63.21	–	38.55	68.34	86.14	–
m-LRCN	21.96	42.02	59.47	–	38.41	64.24	83.21	–	24.58	46.28	63.99	–	39.41	68.65	85.62	–
IMSF-Net	23.73	45.48	63.18	–	38.52	68.50	86.36	–	28.18	51.67	68.84	–	45.36	73.56	88.34	C
1D ResNet18	23.88	45.52	63.16	–	35.46	68.21	86.26	–	27.23	50.22	67.26	–	43.26	71.75	87.12	C
Attention-Based CNN-GRU	21.96	42.56	60.13	–	35.82	64.97	84.06	–	27.31	50.86	68.39	–	44.39	73.26	88.03	C
DeepCNN-BiGRU-Attention	23.93	45.56	63.24	–	38.28	68.30	86.21	–	30.73	54.85	71.37	–	48.08	75.60	88.99	C
STResNet	23.86	45.82	63.54	–	38.93	68.95	86.61	–	33.26	58.86	75.57	–	52.19	79.79	91.62	B

Calibration-free

PP-Net	21.18	40.58	58.03	–	33.88	63.48	82.70	–	21.92	42.90	60.90	–	35.70	65.70	84.27	–
mPP-Net	21.23	40.66	58.07	–	34.16	65.47	82.83	–	22.62	43.51	61.49	–	36.09	65.42	84.09	–
BP-CRNN	22.05	42.29	60.22	–	36.62	65.62	84.17	–	22.47	43.28	60.78	–	36.54	65.78	84.17	–
m-LRCN	21.70	41.68	59.34	–	34.37	64.13	83.33	–	21.94	42.34	59.55	–	34.67	62.93	81.21	–
IMSF-Net	22.42	43.89	62.08	–	37.04	66.10	84.45	–	21.34	40.85	57.42	–	33.15	60.49	79.44	–
1D ResNet18	22.46	43.57	61.08	–	36.92	66.18	84.26	–	20.52	39.73	56.21	–	32.42	59.24	78.04	–
Attention-Based CNN-GRU	22.00	42.21	59.72	–	35.53	64.97	83.60	–	21.40	41.27	58.16	–	33.54	60.87	79.48	–
DeepCNN-BiGRU-Attention	22.89	44.07	61.96	–	36.82	66.23	84.40	–	20.79	39.74	55.97	–	32.64	59.62	78.39	–
STResNet	18.95	36.83	53.52	–	31.45	59.09	79.07	–	21.59	41.40	57.99	–	33.15	60.26	79.01	–
	≥ 60	≥ 85	≥ 95	A	≥ 60	≥ 85	≥ 95	A	≥ 60	≥ 85	≥ 95	A	≥ 60	≥ 85	≥ 95	A
BHS	≥ 50	≥ 75	≥ 90	B	≥ 50	≥ 75	≥ 90	B	≥ 50	≥ 75	≥ 90	B	≥ 50	≥ 75	≥ 90	B
	≥ 40	≥ 65	≥ 85	C	≥ 40	≥ 65	≥ 85	C	≥ 40	≥ 65	≥ 85	C	≥ 40	≥ 65	≥ 85	C

[8] E. O'Brien et al., "The British Hypertension Society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems," J. Hypertens., vol. 8, no. 7, pp. 607–619, 1990.