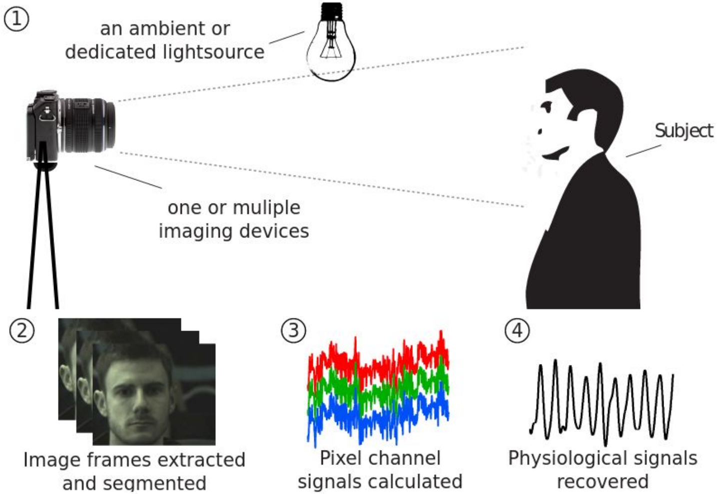
Practice lecture
Vital Signs Estimation using Machine Learning and Deep Learning

Introduction-1

Traditional way to check BP

Modern way to check BP



Related work-1

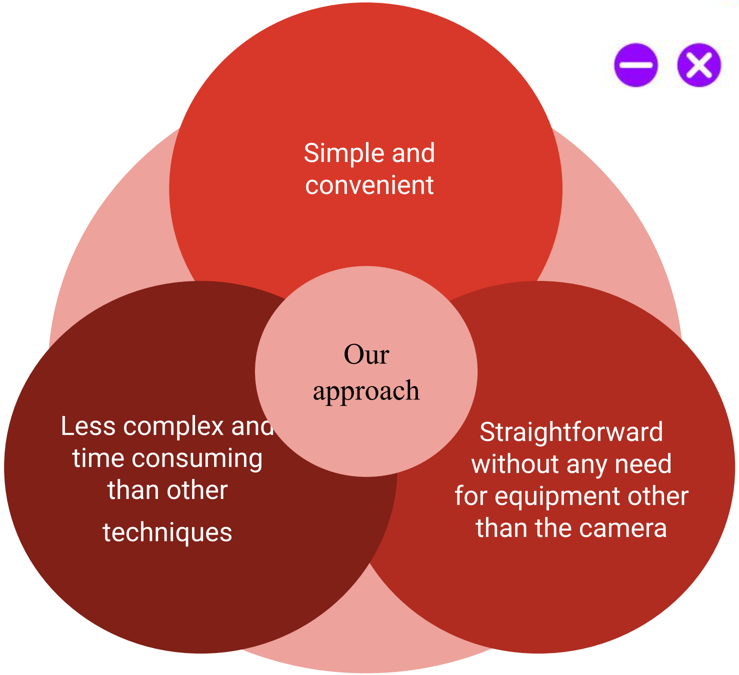
Method	Technique	Results
Luo et al.	Transdermal optical imaging (TOI)	Mean error + SD of 0.39 ± 7.30 mmHg for SBP and −0.20 ± 6.00 mmHg for DBP
Jain et al.	Principal component analysis (PCA)	MAE of 3.9 mmHg and 3.7 mmHg for SBP and DBP, respectively
Secerbegovic et al.	Independent component analysis (ICA)	MAE of 9.48 mmHg for SBP
Oiwa et al	Independent component analysis (ICA)	MAE in range of 1.5–4.5 mmHg
Gaurav et al	Artificial Neural Network	MAE of 4.47 mmHg for SBP and 3.21 mmHg for DBP

Related work-2

Method	Technique	Results
Iuchi et al	Spatial description of the subject's face by extracting time–space information of pulse waves on the face.	MAEs of 6.7 mmHg and 5.4 mmHg, respectively
Schrumpf et al.*	CNN such as AlexNet, Resnet, architecture from another research RNN such as LSTM	ResNet model achieved the lowest SBP MAE of 13.02 mmHg, while AlexNet had the lowest DBP MAE of 8.27 mmHg
Jeong et al.	They studied correlation between blood pressure and image-based pulse transit time (iPTT)	-
Visvanathan et al.	Linear Regression and SVM algorithms	Accuracies of 100% and 99.29% for SBP and DBP, respectively

Introduction to the approach

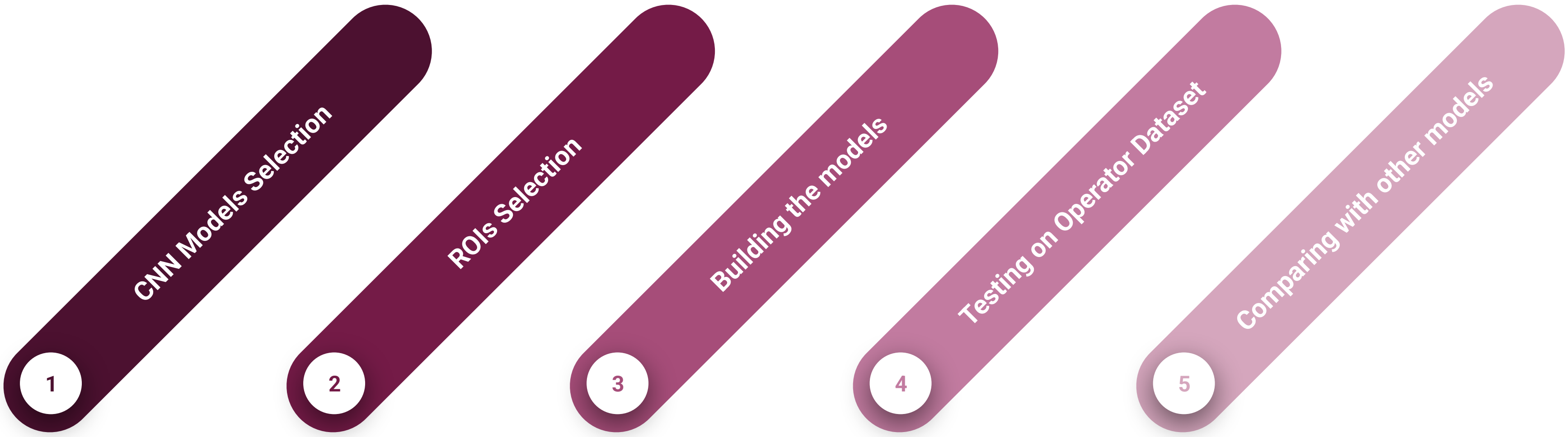
Our approach is based on cropping regions of interest (ROIs), and feeding them directly into a convolutional neural network (CNN), followed by long short-term memory (LSTM) We used LSTM to learn how the changes in the intensities throughout the recording duration lead us to estimate systolic and diastolic blood pressure (SBP and DBP, respectively).



Datasets

- V4V Dataset:
  - 25 frames per second
  - Videos are up to 2 minutes
  - 140 videos were used (divided into 70% training, 20% validation, and 10% testing)
- Operators Dataset:
  - 58 videos have frame rate of 30 fps, whereas one video has 7 fps, and another one has a frame rate of 15 fps
  - The ground truth of respiratory rate of the subject is available for each minute.

Proposed Approach



Dataset pre processing - 1

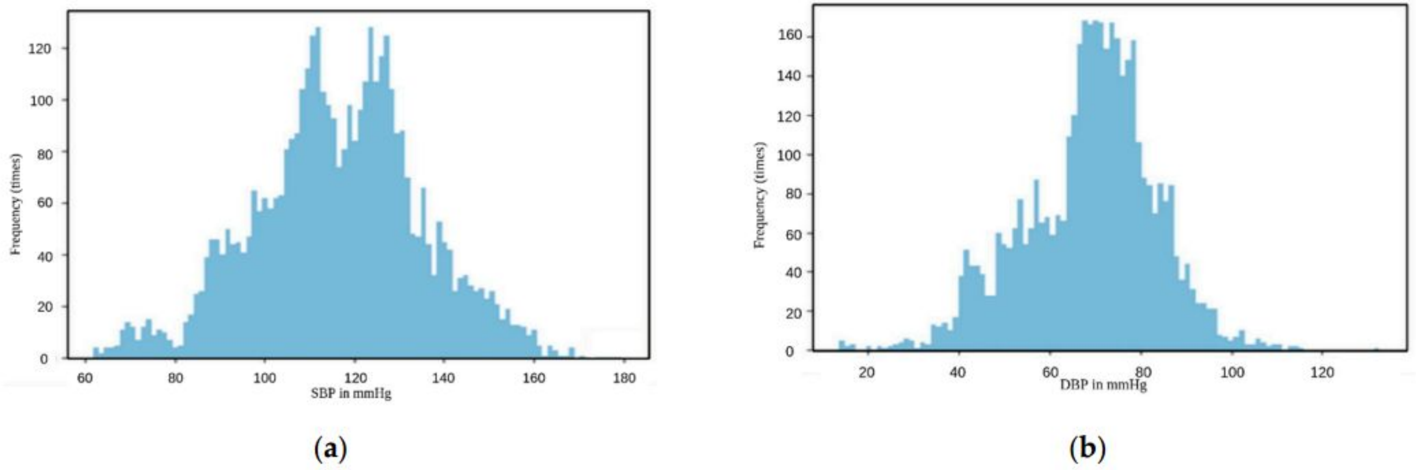
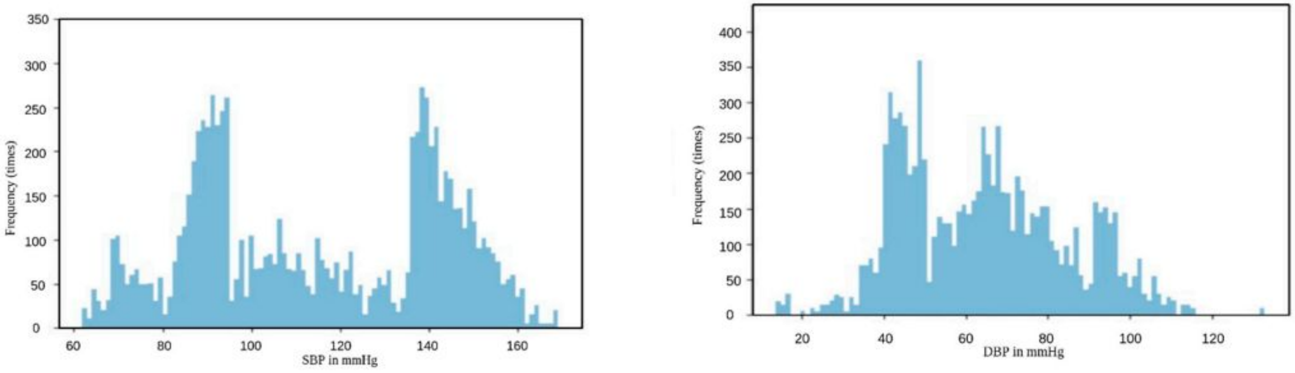


Figure 1. Distribution of the samples before upsampling: (**a**) distribution of SBP samples before upsampling; (**b**) distribution of DBP samples before upsampling.

Dataset pre processing - 2



(a) (b)

Figure 2. Distribution of the samples in the training sets after upsampling: (**a**) distribution of samples used to train models to estimate SBP; (**b**) distribution of samples used to train models to estimate DBP.

CNN Models Selection - 1

- The used models in this project are:
  - Xception
  - DenseNet121
  - VGG16
  - Resnet50V2
  - InceptionV3
  - EfficientNet (B0, B1, B2, B3, B4, B5)

CNN Models Selection - 2

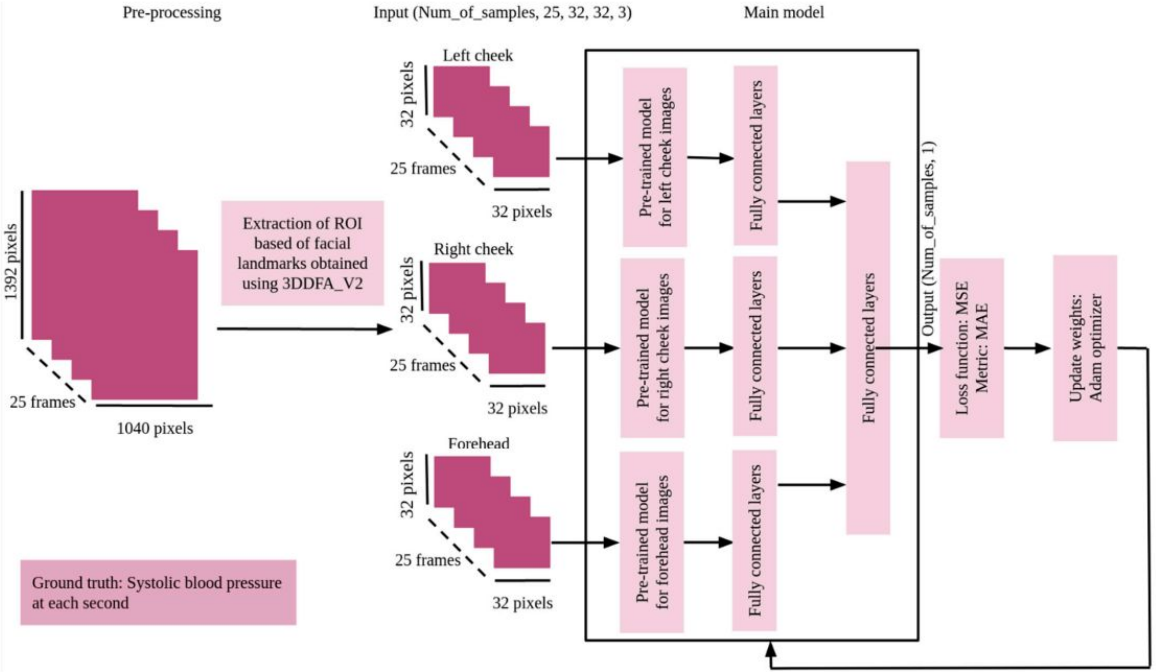


Figure 3. General scheme of the process of choosing CNN models.

CNN Models Selection - 3

Table 1. Result of training different pretrained models.

Pretrained Model	RMSE	MAE
Xception	24.325	21.679
VGG16	21.142	17.178
Resnet50V2	17.427	14.883
InceptionV3	22.648	18.953
DenseNet121	22.358	18.267
EfficientNetB0	16.652	14.237
EfficientNetB1	16.333	14.006
EfficientNetB2	15.958	13.386

Resnet50V2 and EfficientNet (B0, B1, and B2) had the best performance among other pre-trained models (lowest Root Mean Square Error (RMSE) and MAE).

ROI Selection

Three identical models with EfficientNetB3 as a feature extractor were created, each one of them was fed with only one of the regions of interest consecutive images to find which ROI has the most significant impact on the accuracy of the predictions in order to reduce the complexity of our models

Table 2. Result of training EfficientNetB3 with 1 ROI as an input.

ROI	RMSE	MAE
Forehead	24.252	21.218
Left cheek	17.884	15.063
Right cheek	16.332	14.773

Building the models

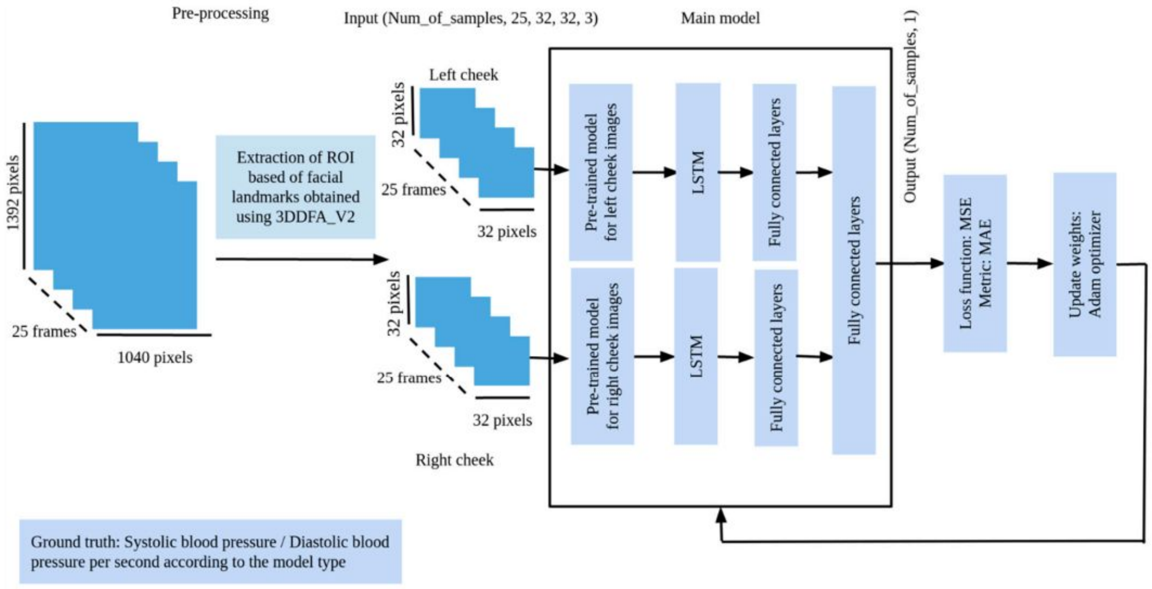


Figure 4. The overall structure of the training process.

Testing on Operators Dataset

We tested our models over the 60 videos of the Operators dataset, obtaining the value of SBP or DBP at every second.

The values that present the SBP or DBP for each minute were calculated by averaging every consecutive 60 values predicted by our models, and then Pearson correlation coefficients which measure the correlation between the predictions of our models and the respiratory rate

Comparing with Other Models

- The extraction of the signal from the videos:
  - Implementation of the POS algorithm
  - Breaking the signal down into intervals.
  - Testing the SNR.
  - Resampling and normalization to zero mean and unit variance

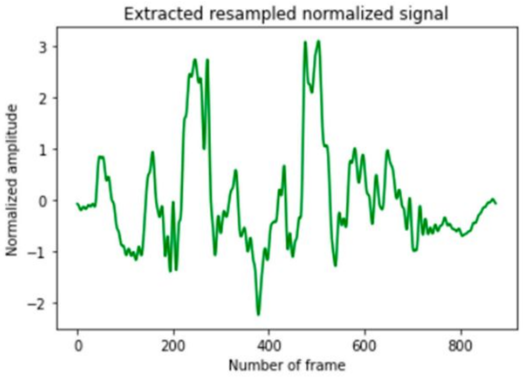


Figure 5. An example of extracted normalized signal from a random 7 s video using POS algorithm.

Results

Table 3. Best models based on the lowest MAE and highest mean accuracy using "V4V test set 1".

Model Type	Number of the Model	MAE	Mean Accuracy
SBP	Model_1	12.408	89.1%
	Model_2	14.491	87.3%
	Model_3	14.174	87.8%
DBP	Model_1	12.652	83.4%
	Model_2	11.371	85.2%
	Model_3	16.071	79.5%

Table 4. Best combinations based on the lowest MAE and highest mean accuracy.

Combination Type	Included Models	MAE	Mean Accuracy (%)
SBP	Model_1, Model_2	11.867	89.5%
	Model_1, Model_3	11.976	89.6%
DBP	Model_1, Model 2	10.706	86.2%
	Model_1, Model_3	12.499	84.1%

Results

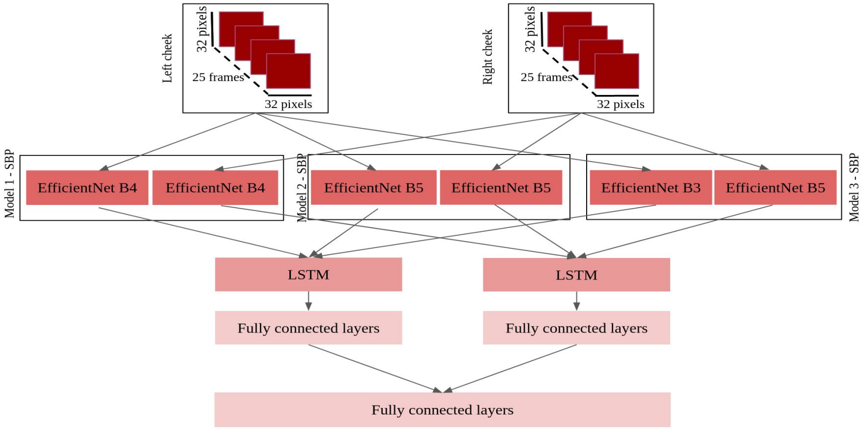
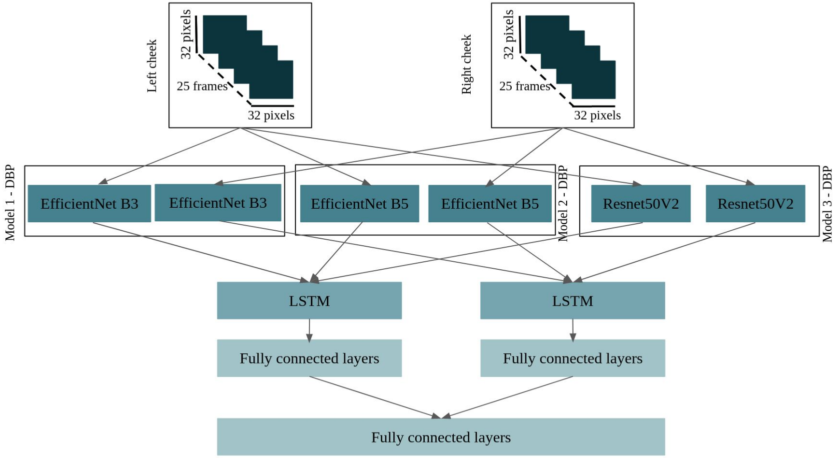


Figure 6. Architecture of the best three SBP models.

Figure 7. Architecture of the best three DBP models.



Results

Table 5. Comparison between our models and available pretrained models using "V4V test set 2".

Model Type	Number of the Model	MAE	Mean Accuracy (%)
SBP	AlexNet [16]	18.582	81.8%
	ResNet [17]	15.456	85.4%
	LSTM	14.579	86.7%
	Slapnicar et al. [18]	31.359	64.3%
	Combination (Model_1, Model 2)	17.477	84.7%
	Combination (Model_1, Model 3)	15.121	86.7%
	Model_1	18.084	83.7%
	Model_2	16.424	84.4%
	Model_3	13.749	88.2%
DBP	AlexNet [16]	13.460	83.1%
	ResNet [17]	12.528	83.9%
	LSTM	13.119	83.5%
	Slapnicar et al. [18]	20.621	62.1%
	Combination (Model_1, Model 2)	12.012	84.04%
	Combination (Model_1, Model 3)	11.169	84.9%
	Model_1	12.391	83.8%
	Model_2	12.697	83.02%
	Model_3	14.833	79.2%

Results

Table 6. Comparison between our models and available pretrained models using Operators dataset.

Model Type	Number of the Model	Average Pearson's Correlation Coefficient over 60 Videos	Average *p*-Value over 60 Videos
SBP	AlexNet [16]	0.4256	0.124
	ResNet [17]	0.4226	0.129
	LSTM	0.4189	0.133
	Slapnicar et al. [18]	-	-
	Combination (Model_1, Model 2)	0.4606	0.107
	Combination (Model_1, Model 3)	0.4863	0.092
	Model_1	0.4670	0.103
	Model_2	0.5028	0.048
	Model_3	0.4304	0.122
DBP	AlexNet [16]	0.4491	0.118
	ResNet [17]	0.4559	0.112
	LSTM	0.3920	0.154
	Slapnicar et al. [18]	-	-
	Combination (Model_1, Model 2)	0.5236	0.045
	Combination (Model_1, Model 3)	0.4623	0.106
	Model_1	0.5302	0.037
	Model_2	0.4743	0.097
	Model_3	0.4429	0.120

Discussion

The challenges:

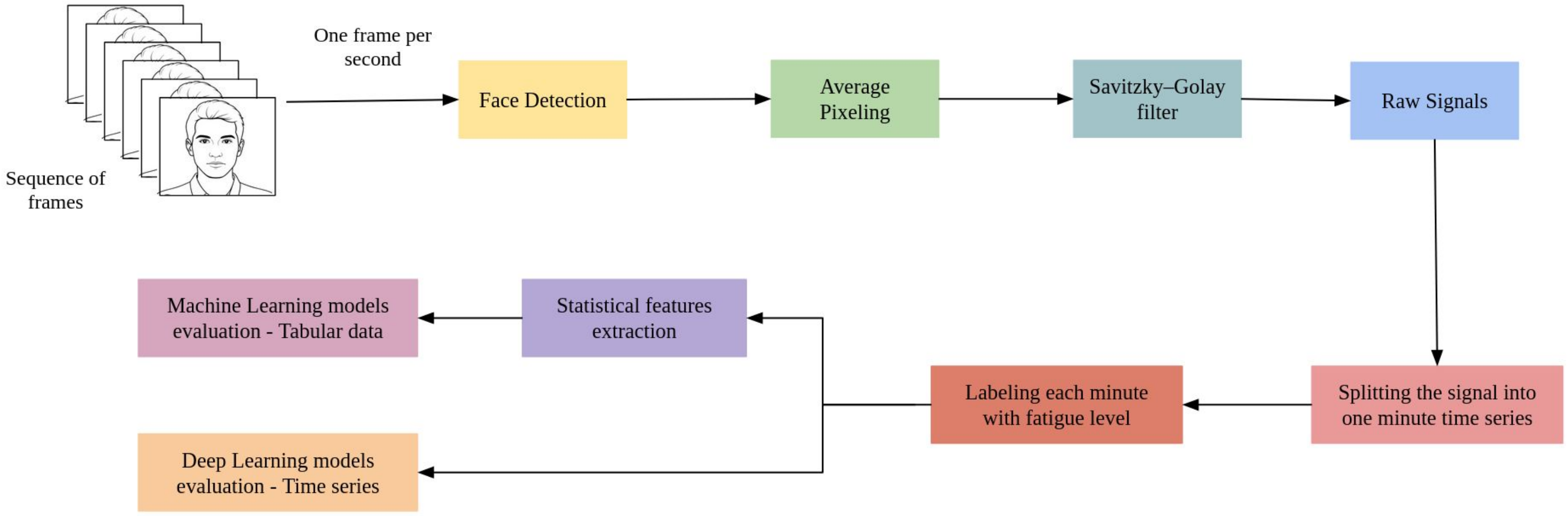
- The dataset has not enough diversity in the subjects' skin color, so our models may not be able to predict the blood pressure of subjects who have dark skin tones accurately.
- The dataset is unbalanced, since there is a lack of unusual values of SBP and DBP.

Future scope:

- Extend the training set with subjects who are older or have darker skin tones with unfamiliar SBP and DBP values due to some medical conditions that the V4V dataset did not cover.
- The approach may be extended to estimate other vital signs such as heart rate, oxygen saturation, body temperature, etc.
- The approach can be easily modified into an application, which will be extremely useful for blood pressure monitoring at home, school, office, or anywhere.

Link to the paper:
https://www.mdpi.com/1424-8220/23/4/1753

Task - 1



One frame per second
Sequence of frames
Face Detection
Average Pixeling
Savitzky–Golay filter
Raw Signals
Splitting the signal into one minute time series
Labeling each minute with fatigue level
Statistical features extraction
Machine Learning models evaluation - Tabular data
Deep Learning models evaluation - Time series

Tasks - 2

Objective: Develop a model to predict mental performance (as an inverse indicator of fatigue) from RGB signal data.

Data: You are provided with RGB signal files from 10 participants. The data is labeled with a mental performance score for each minute.

Task: You may approach this as either:

- A **classification** problem by defining a performance threshold to categorize states (e.g., high vs. low performance), or
- A **regression** problem to predict the continuous performance score directly.

Methodology:

- **Feature Engineering:** Use the raw R, G, B signals. You are encouraged to process these signals and extract meaningful features (e.g., statistical, temporal, frequency-based).
- **Model Input:** Experiment with using 1, 2, or all 3 color channels together to determine the most effective input.
- **Model Exploration:** Investigate a wide range of machine learning and deep learning techniques. The goal is to perform a comparative analysis to identify the optimal model.

Evaluation:

- **For Classifiers:** Accuracy, Precision, Recall, F1-Score, and a Confusion Matrix.
- **For Regressors:** Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

https://drive.google.com/drive/folders/10zPHSi6ZgTNF1adbcWcVdEZkJr7yODR8?usp=sharing

Thank you for your attention!

ITsMOre than a UNIVERSITY