Improving Systolic Blood Pressure Prediction from Remote Photoplethysmography Using a Stacked Ensemble Regressor

L.D. van Putten and K.E. Bamford

Xim Ltd, The University of Southampton Science Park, Southampton

lieke@xim.ai, kate.bamford@xim.ai

Abstract

Hypertension is a serious health risk, and early diagnosis is key to start treatment and avoid fatal complications. We present a stacked ensemble model to predict systolic blood pressure from remote photoplethysmogramy, which enables cuffless measurements. To train the stacked ensemble model, a large dataset with facial remote photoplethysmogramy signals and ground truth values for blood pressure was collected by trained clinicians. From over 4500 measurements 1410 were selected for training following quality control. Over 100 different features were derived from these signals, including statistical features, time domain and frequency domain features. Nine of these features were selected using a forward feature selector. We verified the accuracy of the model on a separately collected validation set. Using a multi-layer perceptron regressor, linear support vector regressor, radial support vector regressor, and ElasticNet for the base models combined with a support vector machine classifier in the stacked ensemble and a RidgeCV model for the final layer, the mean error of the model is reduced to 1.1 mmHg, mean absolute error to 9.5 mmHg and the standard deviation to 12.3 mmHg. Critically, 79% of the hypertensive patients are correctly identified as hypertensive with a prediction over 140 mmHg.

1. Introduction

Untreated hypertension increases the risk of heart disease, strokes and other cardiovascular problems. These types of cardiovascular disease are one of the leading causes of death worldwide [22, 29]. Despite these serious health risks, hypertension often remains undiagnosed as it has no physical symptoms and is therefore referred to as the silent killer [22, 33]. In the US, 1 in 3 adults suffer from hypertension, and of those 1 in 6 are not aware of their condition [26]. Early diagnosis is of major importance to start treatment and avoid fatal complications, however this can only be achieved by actively and accurately monitoring blood

pressure in the population [27].

Traditionally, blood pressure is measured in clinical settings using a sphygmomanometer or more recently using an oscillometric device [32]. These measurements are not a perfect diagnostic tool for hypertension; sphygmomanometers require a skilled clinician to take accurate readings, who can be subjective with regards to the interpretation of Korotkoff sounds and may suffer from terminal digit bias [16, 32]. On top of these observer errors, the accuracy also depends on the placement of the stethoscope as well as on the size and placement of the cuff [16]. The accuracy of oscillometric devices also relies on correct placement and sizing of the cuff and also reduces over time as the cuff should be re-calibrated periodically [9, 16, 23, 32]. The accuracy of both methods can be confounded due to the white coat effect, which causes a temporary spike in a patient's blood pressure. Therefore, after the initial finding of hypertension, patients are often recommended to follow up with ambulatory monitoring or repeated visits to a clinical setting to prove or disprove their diagnosis. Repeated visits to clinical settings are costly, and also inconvenience the patient.

To reduce the cost of blood pressure monitoring and increase the number of people regularly screened for hypertension, research has focused on cuffless blood pressure measurements. One commonly used signal for cuffless blood pressure measurements is photoplethysmography (PPG). PPG signals measure the reflected light from living tissue, which relates to the blood volume present within the optical path [8]. Some cuffless blood pressure measuring methods rely on signals from multiple sensors, for example an electrocardiogram (ECG) and PPG signal [30]. By measuring the time delay between the r-peak in the ECG and systolic peak in the PPG, the pulse transit time can be calculated which is one of the features that can be used to estimate the blood pressure [15]. The pulse transit time can also be estimated when using PPG signals from two different sites [21, 31]. Although blood pressure measuring methods based on the pulse transit time may be able to overcome accuracy errors seen in sphygmomanometers and oscillometric devices, they still rely on specialist equipment

and likely would only be suited to clinical settings. PPG-only measurements, and more specifically single remote-PPG (rPPG) measurements, could open up more accurate blood pressure estimation in clinical and non-clinical settings, without the need for specialist equipment to diagnose hypertension [7].

PPG signals are usually captured using specific equipment with inbuilt light sources and dedicated contact-based sensors, while rPPG signals can be captured with a consumer-grade camera in ambient light as proposed by Verkruysse *et al*. in 2008 [34]. Whilst the ability to use inexpensive cameras is a major advantage in terms of the availability of sensors, it comes with the drawback of increased noise levels due to motion and variable external lighting [21]. Enhanced signal processing algorithms are required to clean the signal before ensemble averaging can be utilised to generate a single representative pulsatile waveform from the rPPG [35]. Blood pressure prediction from a PPG signal from a single site (*e.g*. finger PPG or ear PPG), cannot rely on pulse transit time, because the single waveform does not contain this information. Instead, multiple features are derived from the full rPPG signal and an averaged single pulse waveform is calculated to aid blood pressure prediction. Features derived from the full signal include the heart rate and heart rate variability [10], but can also relate to the average pulse amplitude or symmetry of the signal. Features derived from the ensemble-averaged waveform can include features from the time domain, frequency domain and first and second derivatives of the waveform [8, 14].

These derived features, combined with patient biometrics, have been used in machine learning models to predict blood pressure. Some methods focus on deriving cardiovascular parameters from (r)PPG features such as cardiac output and total peripheral resistance to then estimate the blood pressure [5, 10], while other methods rely purely on the signal's shape to detect features and train a model [3, 6]. These models specifically look at the symmetry of the pulse and the height of the pulse in the diastolic and systolic phase at different percentages of the peak value. Other models derive more complex features from the waveform, its derivative and second derivative that are directly fed into a machine learning model to estimate the blood pressure from the features [15, 21, 30]. In this paper, we use a new dataset consisting of only rPPG signals specifically collected to train machine learning models to predict blood pressure.

2. Model preparation

2.1. Dataset overview

There are several public databases with facial videos for the extraction of rPPG. For example, Bobbia *et al*. published UBFC-RPPG which contained videos of 43 individuals in

2017, [1] and Heusch *et al*. published COHFACE, a different public database with 160 videos in total of 40 individuals [13]. These datasets are useful for testing signal processing and cleaning techniques but do not contain ground truth values for blood pressure, or enough measurements to train a machine learning model. Both are essential to be able to train and test a model that can predict systolic blood pressure from an rPPG signal. The dataset should also contain a wide range of ages and skin tones as well as an even distribution of male and female subjects [14, 19].

Therefore, two datasets were collected specifically for this study. The first dataset contains 1316 subjects, with 4257 measurements in total. This dataset is referred to as the training dataset and was collected in collaboration with hospitals in outpatient appointments [36]. The second dataset was collected purely for validation and contained 86 subjects with 255 measurements in total, chosen to meet the ISO-81060-2:2018 criteria to validate the performance of a blood pressure measuring device compared to an auscultatory reference sphygmomanometer. This dataset is referred to as the validation set.

In the validation set, the ground truth value for blood pressure was measured by two independent observers using an auscultatory sphygmomanometer, while in the training dataset automated oscillometric devices were used. The distribution of ground truths of the two sets is different too, the validation set aims to represent a normal distribution of hypo-, normo- and hypertensive individuals as would be expected in the general population and is required to meet the ISO-81060-2:2018 criteria, while the training set aims to have a more even distribution to aid accurate model training across the full range.

Not all signals collected were of sufficient quality for training and validation. Signals with a signal-to-noise ratio (SNR) below 3.5dB in the frequency domain were excluded [18], as well as signals which had features outside of the anatomically plausible range, indicating an unreliable pulse waveform shape. Following quality control, 1410 signals remained in the training set belonging to 613 subjects, and 183 signals remained in the validation set belonging to 69 subjects. In the training set, 27% of measurements have a ground truth systolic blood pressure below 120 mmHg, 36% between 120 to 140 mmHg, and 37% above 140 mmHg. Of the 613 subjects in the training set, 35% were already receiving treatment for hypertension. No individuals receiving treatment for hypertension were accepted in the validation set.

2.1.1 Feature extraction

As shown in Fig. 1, filtering, normalization and baseline correction are performed before features are extracted from the rPPG signal. This cleaned signal is used directly to cal-
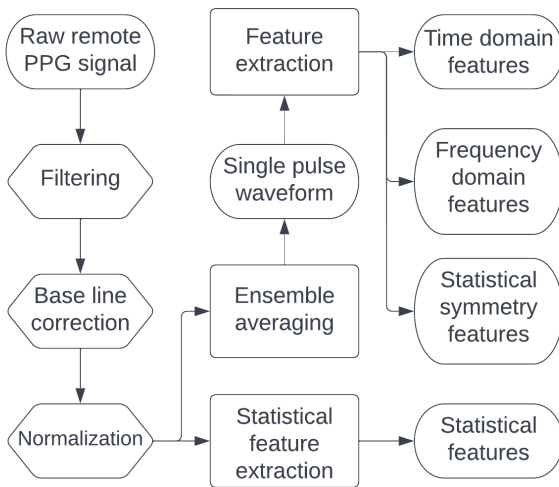
Figure 1. Block diagram showing the feature extraction process.

culate statistical features, and further processed using ensemble averaging to generate a single waveform. This single pulse waveform is used to calculate time-domain features, frequency-domain features as well as further statistical features.

2.2. Model types overview

A wide range of model types has been used to predict blood pressure from features derived from (r)PPG signals. Popular machine learning models include deep learning approaches, [3, 28], linear and non-linear regression models [14, 17, 25] and decision tree based algorithms [17]. Although these models have been shown to find patterns in the rPPG signals that allows them to predict the blood pressure, they all have their own strengths and weaknesses. Decision tree based models are prone to overfitting and may not generalize well. Deep learning approaches require large amounts of data, and do not create human interpretable models. Linear regression methods require features that have a linear correlation to the target value, which isn't always the case for complex physiological features. Non-linear regression methods may struggle to find the optimum function in noisier datasets, which usually is the case for remote PPG signals due to the introduction of environmental and movement disturbances to the signal.

2.2.1 Stacked Ensemble

To combat the weaknesses of individual models, a stacked ensemble model utilises multiple trained and optimised base models as input for a second layer model. The second layer model can learn the strengths and weaknesses

of the individual base models and achieve better accuracy. This can be especially helpful in noisy datasets, because the stacked ensemble can reduce the impact of noisy data by using diverse models, optimising the combination of base models, and weighing the models based on performance in different ranges of the prediction [2, 4, 11].

2.3. Evaluation procedure

Model performance is evaluated on the withheld validation set, where the mean error (ME) and standard deviation (SD) are calculated between the ground truth systolic blood pressure and the systolic pressure predicted by the stacked ensemble. Of the 255 signals collected in the validation set, 183 were of sufficient quality to generate clustered waveforms for feature generation. To enable a fair comparison of the model performance under the criteria as outlined in ISO-81060-2:2018 criteria, a randomly selected subset of the validation set was used that meets the correct normal distribution, as a result, 139 measurements remained in the final validation set.

The performance of the feature based stacked ensemble model will be compared against the performance of a model that is trained using only patient biometrics (including weight). A further metric on top of the ME and SD of each model is the percentage of correctly diagnosed hypertensives. The British Hypertension Society classifies anyone with a systolic blood pressure above 140 mmHg as hypertensive [24], giving a total of 28 hypertensives in the validation set.

3. Model implementation

3.1. Feature selection

From the feature extraction process more than 100 features are extracted, however, some of these features are highly correlated. Highly correlated features can lead to overfitting and decrease model interpretability and portability [12, 20]. On top of these effects on the performance of the model, a higher number of features also increases the computational time needed to train the model [20], so careful feature selection is vital.

On top of the derived features, both the training and validation datasets contain four biometric features which can be used for model training: age, height, sex, and weight. Although weight commonly has a strong correlation with blood pressure, this correlation does not necessarily hold up in acutely unwell patients. In feature selection and training for the feature based model, weight was therefore excluded as it could affect model accuracy when used in clinical settings.

Feature selection was undertaken as the first part of the process workflow, shown in Fig. 2a. Before the features were selected, the biometric features (age, height and sex)
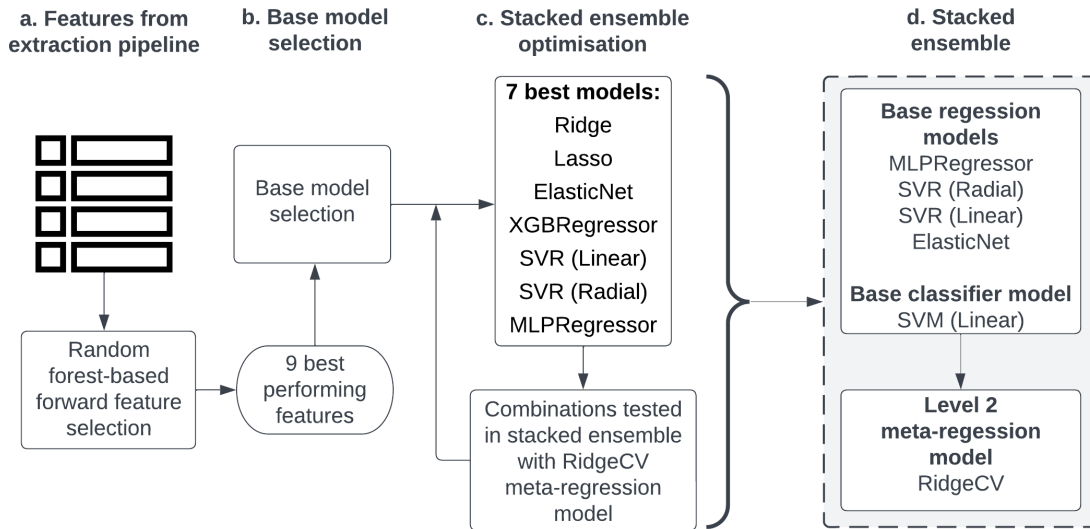
Figure 2. Block diagram showing the process workflow of the stacked ensemble training and optimisation.

were linearly combined into a single combination feature. After this step, forward feature selection was used to select the optimal feature set, using feature importance in a random forest model as the evaluation metric. Nine features were selected: the first feature was the linearly combined biometric feature, four features were time domain features, three features were statistical symmetry features, and one feature was a statistical feature derived from the full signal. These nine best performing features were used across the base model training and optimising.

3.2. Base model selection and optimisation

Multiple regression models across a range of models types were trained and validated with default parameters to identify the most appropriate base models to combine in the stacked ensemble (Fig. 2b). They were chosen to represent a range of supervised regression methods, and included linear models, tree-based methods, support vector machines and neural networks. The chosen models including parameters where different from the default can be seen in Tab. 1. Base models were evaluated using the mean absolute error (MAE).

Stacked ensemble models rely on the combination of diverse base models to reduce bias and generalisation error [2, 4]. In order to select the combination of base models that maximise performance without adding additional unnecessary complexity, the seven best-performing single base models were optimised individually and then tested in combination in an iterative process where each iteration included between four and seven models, shown in Fig. 2c.

Candidate base model	Parameters[1]
Elastic Net	
Gamma	
Lasso	
Ridge	
Random Forest	
Extra-Trees	
XGBoost	
AdaBoost	
Bagging	
Support Vector Regressor (1)	kernel = 'radial'
Support Vector Regressor (2)	kernel = 'linear'
K nearest neighbours	
Multi-layer perceptron (1)	hidden_layer_sizes = (100,50)
Multi-layer perceptron (2)	hidden_layer_sizes = (200,50)

[1] Unless otherwise stated, parameters were the sklearn default parameters.

Table 1. List of candidate models

Previous studies have found that between three and four base learners are often optimal compared to combining all possible base learners in the stacked ensemble [2,37]. Here, seven models were chosen as an acceptable trade-off between computational testing time required to optimise and test all possible stacking combinations and the possible addition of more diverse base models.

Model	MAE (mmHg)	ME (mmHg)	SD (mmHg)
MLP (2)	11.1	3.6	13.9
Linear SVR	11.5	1.0	15.1
Radial SVR	11.7	4.4	14.3
Extra-Trees	12.3	4.4	14.5
Lasso	12.4	5.2	14.8
Ridge	12.5	5.4	14.8
ElasticNet	12.5	5.5	14.9
Random Forest	12.7	4.6	14.9
Bagging	12.8	5.2	15.0
MLP (1)	12.8	5.9	14.9
XGB	13.4	4.3	16.2
KNN	13.5	4.7	16.5
AdaBoost	16.1	10.5	15.7
Gamma	19.0	10.2	19.9

SVR: Support Vector Regressor; MLP (1): Multi-layer perceptron where hidden_layer_size = (100,50); MLP (2): Multi-layer perceptron where hidden_layer_size = (200,50) XGB: XGBoost; KNN: K nearest neighbours

Table 2. Comparison of individual base models prior to hyperparameter optimisation. The best-performing seven models by MAE are listed above the dashed line.

3.3. Stacked ensemble learning framework

The stacking framework consists of a two-layer structure, shown in Fig. 2d. The first layer contains *n* base models and the second layer contains a single meta model. In order to build the stacking framework, the training data was split into five folds, and for each base learner the first four folds were used for training and the remaining folds were used for predictions for each base model. The out-of-fold predictions were then passed as new features to the second layer meta model. The meta model was initially trained on the combined base model predictions and the original selected features.

Finally, a classifier was used to classify the data into three classes ($\leqslant$120 mmHg, 120-140 mmHg, $\geqslant$140 mmHg). In order to select a classifier, a range of model types were individually trained and tested, including random forest, logistic regression, and support vector machines (SVM) using both linear and radial kernels. The model with the highest accuracy was a support vector machine classifier (SVM) with a linear kernel. The probability scores for each of the classes from the classifier are used as an input for the second level model as additional features. The final optimised model is then used to generate predictions for the validation set and to report on the accuracy.

Model	MAE (mmHg)	ME (mmHg)	SD (mmHg)
Biometric only ensemble	14.7	7.7	16.3
Feature-based ensemble	9.5	1.1	12.3

Table 3. Comparison of MAE, ME and SD of the two different models.

4. Results

4.1. Individual model performance

All models were trained and validated following feature selection to allow for direct comparison. MAE, ME and SD of the individual base models is shown in Tab. 2. The base models had a ME between 4.0 and 5.8 and SD ranging between 13.8 and 15.5. The neural network algorithm, a multi-layer perceptron (MLP) had the lowest SD at 13.9 mmHg.

The two support vector regressor (SVR) models were the next best performing base models by MAE. The SVR with a linear kernel had the lowest ME (1.0 mmHg) but this was at the expense of the SD, which is the highest amongst the best performing base models at 15.1 mmHg. The SVR trained with a radial kernel had a lower SD (14.3 mmHg), but the ME remained above 4 mmHg. The two different kernels applied to the data mean that the algorithms handle the linear and non-linear data differently, providing a diverse set of models despite their similarities.

Of the four linear models tested as candidate models, three were in the best-performing base model selection. All three models are forms of regularised linear algorithms, assigning penalties to coefficients in order to reduce overfitting. However, they all had ME values over 5 mmHg, showing none of them would perform well enough to predict systolic blood pressure as a single model.

Following iterative testing of stacked ensemble models containing between four and seven of the best-performing base models, the model with the best performance was a combination of four base models: SVR (radial), SVR (linear), Elastic Net and MLP. Finally, the SVM classifier was tested and the probabilities for each class were added as features. The classifier achieves an accuracy of 64% in the validation data.

4.2. Stacked ensemble

To show the rPPG features provide superior information when predicting systolic blood pressure, the results of a model trained only on the collected biometric features (height, weight, age and sex) were compared to the results of the final feature based model (Tab. 3). The rPPG feature-based model outperformed the biometric model in all evaluated metrics, with a reduction in MAE of more than 5
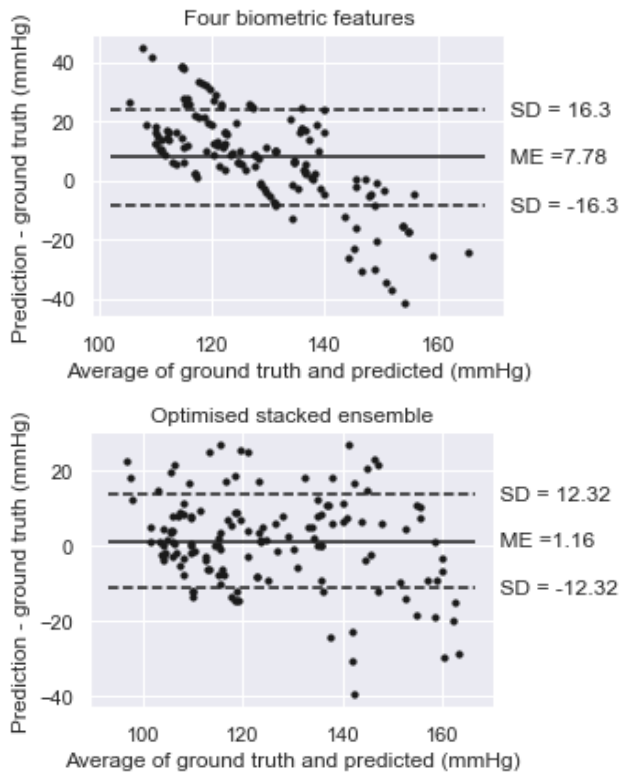
Figure 3. Bland-Altman plot for stacked ensemble model using four biometric features (top) and final feature based stacked ensemble model (bottom).
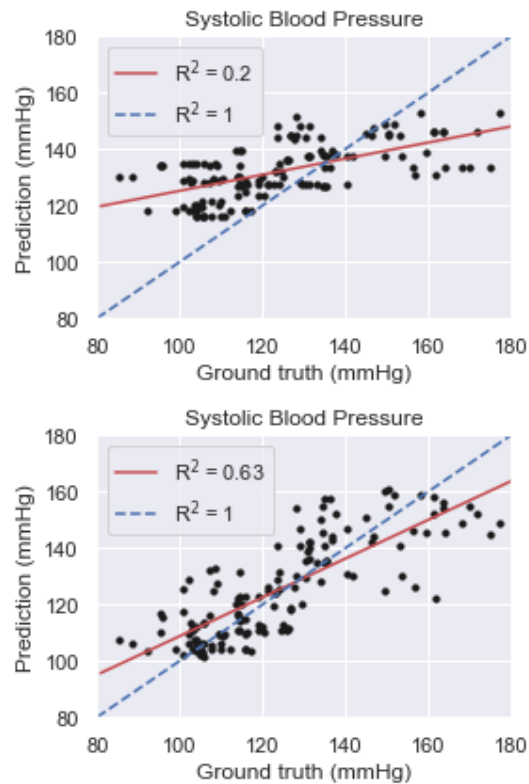
Figure 4. Regression plot comparing prediction and ground truth for the model using only biometric features (top) and the feature based stacked ensemble model (bottom). Red solid line indicates the R^2 between ground truth and prediction, blue dashed line indicates the ideal correlation.

mmHg, in ME more than 6 mmHg, and 4 mmHg in SD.

The Bland-Altman plots for each model are shown in Fig. 3. The biometric model has a clear proportional bias, where samples in the hypertensive range ($\geqslant$140 mmHg) are under-predicted and samples in the hypotensive range ($\leqslant$120 mmHg) are over-predicted. Just over half (54%) of the hypertensive individuals were correctly provided a value over 140 mmHg. In comparison, the stacked ensemble model including the rPPG features (Fig. 3) predicted 79% of the samples in the hypertensive range correctly as hypertensive.

Fig. 4 shows the difference between the predicted values and the ground truth for both stacked ensemble models. The R^2 value for the biometric only model is significantly lower than the feature based stacked ensemble at 0.2 compared to 0.63, showing a closer match to the predicted values in the feature based model. The biometric only model also has predictions limited between 116 and 153 mmHg, despite the systolic blood pressure in the training data ranging between 84 and 219 mmHg. This tendency towards predicting the mean can be attributed to noisy data, because the error in the independent variable causes linear regression slopes to tend towards zero. Although both models show this effect

to some extent, it is much less pronounced in the feature based model, also indicated by the higher R^2 value.

The stacked ensemble combination also outperforms the individual base models, with a reduction in MAE and SD compared to all four base models. The ME is significantly lower compared to three of the four base models, and not significantly different from the linear SVR. These results show that the meta-regression model has successfully learned to take the output of each of the individual models to create a more accurate prediction.

5. Conclusion

In this paper, we presented a stacked ensemble model that utilises features extracted from facial rPPG signal to estimate the systolic blood pressure. The stacked ensemble model was trained on a clinically collected dataset of videos containing participants with a wide range of blood pressures. This model was shown to effectively estimate the systolic blood pressure in the validation set with a mean error of 1.1 mmHg and standard deviation of 12.3 mmHg

compared to the ground truth values taken using a sphygmomanometer. Using a stacked ensemble model, the ME reduced significantly (by 2.5 mmHg compared to the MLP model alone). The SD of the stacked ensemble is 2.6 mmHg lower compared to the ElasticNet model, which has the largest SD, and 1.6 mmHg lower compared to the MLP, which has the smallest SD of all the models in the stacked ensemble. Overall, combining the different base models into the stacked ensemble has been shown to improve the performance of the final prediction.

The stacked ensemble not only performs better than the individual models in the stacked ensemble, but also outperforms a model that relies solely on biometric features as shown in Tab. 3. The stacked ensemble correctly diagnoses 79% of the hypertensive subjects as hypertensive, while the biometric only model only achieves 54%. This shows that the stacked ensemble prediction from a facial rPPG signal could be a helpful addition for clinicians providing virtual healthcare services, by adding a more reliable prediction of the blood pressure than a clinician could plausibly make based on patient biometrics alone in a remote healthcare setting.

In our future work, we aim to explore the rPPG signal in more detail in order to find more independent features and work on improved methods of signal cleaning to increase the SNR in the collected signals. An increased SNR will increase the size of our usable signal database and combined with novel features this could improve the accuracy of a trained stacked ensemble even further.

References

[1] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 124:82–90, 2019. 2

[2] Leo Breiman. Stacked regressions. *Machine learning*, 24:49–64, 1996. 3, 4

[3] Benedetta C Casadei, Alessandro Gumiero, Giorgio Tantillo, Luigi Della Torre, and Gabriella Olmo. Systolic blood pressure estimation from ppg signal using ann. *Electronics*, 11(18):2909, 2022. 2, 3

[4] Angelos Chatzimparmpas, Rafael M. Martins, Kostiantyn Kucher, and Andreas Kerren. Stackgenvis: Alignment of data, algorithms, and models for stacking ensemble learning using performance metrics. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1547–1557, 2021. 3, 4

[5] Anirban Dutta Choudhury, Rohan Banerjee, Aniruddha Sinha, and Shaswati Kundu. Estimating blood pressure using windkessel model on photoplethysmogram. In *2014 36th annual international conference of the IEEE engineering in medicine and biology society*, pages 4567–4570. IEEE, 2014. 2

[6] Pedro Henrique de Brito Souza, Israel Machado Brito Souza, Symone Gomes Soares Alcalá, Priscila Valverde

de Oliveira Vitorino, Adson Ferreira da Rocha, and Talles Marcelo Gonçalves de Andrade Barbosa. Video-based photoplethysmography and machine learning algorithms to achieve pulse wave velocity. *International Journal of Biotech Trends and Technology*, 11(1):7–15, 2021. 2

[7] Mohamed Elgendi. On the analysis of fingertip photoplethysmogram signals. *Current cardiology reviews*, 8(1):14–25, 2012. 2

[8] Mohamed Elgendi, Richard Fletcher, Yongbo Liang, Newton Howard, Nigel H Lovell, Derek Abbott, Kenneth Lim, and Rabab Ward. The use of photoplethysmography for assessing hypertension. *NPJ digital medicine*, 2(1):60, 2019. 1, 2

[9] Merrill F Elias and Amanda L Goodell. Human errors in automated office blood pressure measurement: still room for improvement. *Hypertension*, 77(1):6–15, 2021. 1

[10] Hayato Fukushima, Haruki Kawanaka, Md Shoaib Bhuiyan, and Koji Oguri. Cuffless blood pressure estimation using only photoplethysmography based on cardiovascular parameters. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2132–2135. IEEE, 2013. 2

[11] M.A. Ganaie, Minghui Hu, A.K. Malik, M. Tanveer, and P.N. Suganthan. Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115:105151, 2022. 3

[12] Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004. 3

[13] Guillaume Heusch, André Anjos, and Sébastien Marcel. A reproducible study on remote heart rate measurement. *arXiv preprint arXiv:1709.00962*, 2017. 2

[14] Monika Jain, Sujay Deb, and A Venkata Subramanyam. Face video based touchless blood pressure and heart rate estimation. In *2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–5. IEEE, 2016. 2, 3

[15] Da Un Jeong and Ki Moo Lim. Combined deep cnn–lstm network-based multitasking learning architecture for noninvasive continuous blood pressure estimation using difference in ecg-ppg features. *Scientific Reports*, 11(1):13539, 2021. 1, 2

[16] Noa Kallioinen, Andrew Hill, Mark S Horswill, Helen E Ward, and Marcus O Watson. Sources of inaccuracy in the measurement of adult patients' resting blood pressure in clinical settings: a systematic review. *Journal of hypertension*, 35(3):421, 2017. 1

[17] Syed Ghufran Khalid, Jufen Zhang, Fei Chen, Dingchang Zheng, et al. Blood pressure estimation using photoplethysmography only: comparison between different machine learning approaches. *Journal of healthcare engineering*, 2018, 2018. 3

[18] Adam Kiddle, Helen Barham, Simon Wegerif, and Connie Petronzio. Dynamic region of interest selection in remote photoplethysmography: Proof-of-concept study. *JMIR Formative Research*, 7:e44575, 2023. 2

[19] Mayank Kumar, Ashok Veeraraghavan, and Ashutosh Sabharwal. Distanceppg: Robust non-contact vital signs moni-

toring using a camera. *Biomedical optics express*, 6(5):1565–1588, 2015. 2

[20] Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P Trevino, Jiliang Tang, and Huan Liu. Feature selection: A data perspective. *ACM computing surveys (CSUR)*, 50(6):1–45, 2017. 3

[21] Ye Lu, Chaoqun Wang, and Max Q-H Meng. Video-based contactless blood pressure estimation: A review. In *2020 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, pages 62–67. IEEE, 2020. 1, 2

[22] Katherine T Mills, Andrei Stefanescu, and Jiang He. The global epidemiology of hypertension. *Nature Reviews Nephrology*, 16(4):223–237, 2020. 1

[23] Martin G Myers, Roland Asmar, and Jan A Staessen. Office blood pressure measurement in the 21st century. *The Journal of Clinical Hypertension*, 20(7):1104–1107, 2018. 1

[24] Eoin O'Brien, James Petrie, William Littler, Michael de Swiet, Paul L Padfield, Kevin O'Malley, Michael Jamieson, Douglas Altman, Martin Bland, and Neil Atkins. The british hypertension society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems. *Journal of hypertension*, 8(7):607–619, 1990. 3

[25] Kosuke Oiwa, Shizuka Bando, and Akio Nozawa. Contactless blood pressure sensing using facial visible and thermal images. *Artificial Life and Robotics*, 23:387–394, 2018. 3

[26] Soyoun Park, Cathleen Gillespie, Jason Baumgardner, Quanhe Yang, Amy L Valderrama, Jing Fang, Fleetwood Loustalot, and Yuling Hong. Modeled state-level estimates of hypertension prevalence and undiagnosed hypertension among us adults during 2013-2015. *The Journal of Clinical Hypertension*, 20(10):1395–1410, 2018. 1

[27] Thomas G Pickering, John E Hall, Lawrence J Appel, Bonita E Falkner, John Graves, Martha N Hill, Daniel W Jones, Theodore Kurtz, Sheldon G Sheps, and Edward J Roccella. Recommendations for blood pressure measurement in humans and experimental animals: part 1: blood pressure measurement in humans: a statement for professionals from the subcommittee of professional and public education of the american heart association council on high blood pressure research. *Hypertension*, 45(1):142–161, 2005. 1

[28] KNG Priyanka, Paul C-P Chao, Tse-Yi Tu, Yung-Hua Kao, Ming-Hua Yeh, Rajeev Pandey, and Fitrah P Eka. Estimating blood pressure via artificial neural networks based on measured photoplethysmography waveforms. In *2018 IEEE SENSORS*, pages 1–4. IEEE, 2018. 3

[29] Gregory A Roth, Degu Abate, Kalkidan Hassen Abate, Solomon M Abay, Cristiana Abbafati, Nooshin Abbasi, Hedayat Abbastabar, Foad Abd-Allah, Jemal Abdela, Ahmed Abdelalim, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet*, 392(10159):1736–1788, 2018. 1

[30] Fabian Schrumpf, Patrick Frenzel, Christoph Aust, Georg Osterhoff, and Mirco Fuchs. Assessment of non-invasive blood pressure prediction from ppg and rppg signals using deep learning. *Sensors*, 21(18):6022, 2021. 1, 2

[31] Manuja Sharma, Karinne Barbosa, Victor Ho, Devon Griggs, Tadesse Ghirmai, Sandeep K Krishnan, Tzung K Hsiai, Jung-Chih Chiao, and Hung Cao. Cuff-less and continuous blood pressure monitoring: a methodological review. *Technologies*, 5(2):21, 2017. 1

[32] Saulat Siddique, Aamir Hameed Khan, Hunaina Shahab, Yu-Qing Zhang, Jam Chin Tay, Peera Buranakitjaroen, Yuda Turana, Narsingh Verma, Chen-Huan Chen, Hao-Min Cheng, et al. Office blood pressure measurement: A comprehensive review. *The Journal of Clinical Hypertension*, 23(3):440–449, 2021. 1

[33] Vamsi Krishna Undavalli and H Mp. Prevalence of undiagnosed hypertension: a public health challenge. *Int J Community Med Public Health*, 5(4):1366–1370, 2018. 1

[34] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008. 2

[35] William Waugh, John Allen, James Wightman, Andrew J Sims, and Thomas AW Beale. Novel signal noise reduction method through cluster analysis, applied to photoplethysmography. *Computational and mathematical methods in medicine*, 2018, 2018. 2

[36] Laura Wiffen, Thomas Brown, Annika Brogaard Maczka, Melissa Kapoor, Laurence Pearce, Milan Chauhan, Anoop J Chauhan, Manish Saxena, Lifelight Trials Group, et al. Measurement of vital signs by Lifelight software in comparison to standard of care multisite development (vision-md): Protocol for an observational study. *JMIR Research Protocols*, 12(1):e41533, 2023. 2

[37] Yuzhen Zhang, Jun Ma, Shunlin Liang, Xisheng Li, and Jindong Liu. A stacking ensemble algorithm for improving the biases of forest aboveground biomass estimations from multiple remotely sensed datasets. *GIScience & Remote Sensing*, 59(1):234–249, 2022. 4