Development and validation of a deep learning-based automatic auscultatory blood pressure measurement method

Pan, F., He, P., Wang, H., Xu, Y., Pu, X., Zhao, Q., Chen, F. & Zheng, D

DOI 10.1016/j.bspc.2021.102742
ISSN 1746-8094

Publisher: Elsevier

Development and Validation of a Deep Learning-based Automatic Auscultatory Blood Pressure Measurement Method

Fan Pan [a], Peiyu He [a], He Wang [b], Yuhang Xu [c], Xiaobo Pu [d], Qijun Zhao [e], Fei Chen*[f] and Dingchang Zheng*[c]

[a] College of Electronics and Information Engineering, Sichuan University, Chengdu 610064, China

[b] School of Computing, University of Leeds, Leeds LS2 9JT, UK

[c] Research Centre of Intelligent Healthcare, Faculty of Health and Life Science, Coventry University, Coventry CV1 5FB, UK

[d] Department of Cardiology, West China Hospital, Sichuan University, Chengdu 610041, China

[e] College of Computer Science, Sichuan University, Chengdu 610064, China

[f] Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen 518055, China

*Address correspondence to Fei Chen, Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen 518055, China. Electronic mail: fchen@sustech.edu.cn

*Address correspondence to Dingchang Zheng, Research Centre of Intelligent Healthcare, Faculty of Health and Life Science, Coventry University, Coventry CV1 5FB, UK. Electronic mail: ad4291@coventry.ac.uk

Abstract

Objective: Manual auscultatory is the gold standard for clinical non-invasive blood pressure (BP) measurement, but its usage is decreasing as it requires much professional skills and training, and its environmental concerns related to mercury toxicity. As an alternative, automatic oscillometric technique has been used as one of the most common methods for BP measurement, however, it only estimates BPs based on empirical equations. To overcome these problems, this study aimed to develop a deep learning-based automatic auscultatory BP measurement method, and clinically validate its performance.

Methods: A deep learning-based method that utilized time-frequency characteristics and temporal dependence of segmented Korotkoff sound (KorS) signals and employed convolutional neural network (CNN) and long short-term memory (LSTM) network was developed and trained using KorS and cuff pressure signals recorded from 314 subjects. The BPs determined by the manual auscultatory method was used as the reference for each measurement. The measurement error and BP category classification performance of our proposed method were then validated on a separate dataset of 114 subjects. Its performance in comparison with the oscillometric method was also comprehensively analyzed.

Results: The deep learning method achieved measurement errors of 0.2 ± 4.6 mmHg and 0.1 ± 3.2 mmHg for systolic BP (SBP) and diastolic BP (DBP), respectively, and achieved high sensitivity, specificity and accuracy (all > 90%) in classifying hypertensive subjects, which was better than those of oscillometric method.

Conclusion: This validation study demonstrated that deep learning-based automatic auscultatory BP measurement can be developed to achieve high measurement accuracy and high BP category classification performance.

Keywords: Blood pressure measurement, deep learning, manual auscultatory method, oscillometric method.

Introduction

High blood pressure (BP) is one of the major modifiable risk factors for cardiovascular disease, aggravating the greatest global burden of disease.[1, 2] Accurate BP measurement helps to identify the presence of high BP, which contributes to reducing the risk of future cardiovascular events, whereas, inaccurate BP estimation may result in serious clinical consequences.[3, 4] Even 5 mmHg error either above or below the actual BP would result in tens of million people being exposed to unnecessary treatment or being denied treatment.[5]

Manual auscultatory method is the gold standard for non-invasive clinical BP measurement.[6] This method auscultates the brachial artery with a stethoscope detecting the appearance as well as the disappearance or muffling of the Korotkoff sounds (KorS), which corresponds to the systolic blood pressure (SBP) and diastolic blood pressure (DBP), respectively.[7] However, since the manual auscultatory method always use a mercury manometer for measuring the pressure, and requires professional skills and training, it has been widely replaced by automatic technique in clinic, ambulatory, home and hospital settings.[8] The oscillometry is one of the most common techniques for automatic BP measurement. In principle, the automatic oscillometric method estimates mean arterial BP (MAP) from the cuff pressure when the oscillation amplitude is maximal, and then mathematically computes SBP and DBP using empirical ratios derived from the recorded oscillometric waveform envelope.[8, 9] Since the empirical ratios are obtained based on population averages during device development, the BPs

determined by oscillometric technique are actually ‘empirical’ BPs. Therefore, automatic oscillometric method only estimates BPs, and does not actually measure BPs in theory. Furthermore, it is difficult for the oscillometric technique to accurately estimate BPs in hypertensive subjects and in subjects with arterial stiffness, which are common conditions that occur with aging and many diseases.[10, 11] Therefore, investigating an alternative method which can achieve accurate BP measurement is still clinically important.[12]

For the gold standard manual auscultatory method, accurate identification of the KorS features is the key to detect the appearance and disappearance of the KorS, from which BPs are measured. It is therefore crucial for automatic methods to be able to identify these features. However, due to the poor ability of traditional signal processing technologies (e.g., time-frequency analysis or power spectrum analysis) to recognize complex features, it is difficult to accurately differentiate the KorS features (e.g., amplitude, frequency range or during time) between subjects to aid BP determination. To the best of our knowledge, there is still no well-accepted clinic automatic auscultatory BP monitor that has been widely adapted in the healthcare system.

Unlike traditional signal processing technology, deep learning technique has multiple layers of non-linear processing and do not rely on feature selection to obtain reliable results. Deep learning techniques have been widely applied to a variety of medical fields, such as diabetic retinopathy detection[13], arrhythmia detection and classification[14], hypertrophic cardiomyopathy detection[15], and disease prediction[16], where impressive outcomes have been achieved. We have recently developed a new

deep learning-based automatic auscultatory BP measurement method in a preliminary study and evaluated its performance under different measurement conditions (resting, deeper breathing, talking and arm movement) with normotensive subjects.[17, 18] These investigations have demonstrated the significant potential of using deep learning technique to automatically measure BP accurately. However, its performance has not been clinically validated on subjects with a wide range of BPs, and also has not been compared to the commonly used automatic oscillometric BP measurement method. This study aims to provide clinical evidence on the measurement accuracy as well as on the performance of classifying different BP categories.

This paper is organized as follows. In the second part of this paper, we introduce a deep learning-based automatic BP measurement method employed CNN and LSTM. The third part includes the results of measurement accuracy of the deep learning method and the comparison of its performance with oscillometric method. The fourth part discusses and analyses the results, and displays the advantages of our method in comparison with oscillometric method. Finally, the fifth part concludes the paper.

Methods

Figure 1 shows the overall methodology flow of this study. The KorS and cuff pressure signals were simultaneously recorded during reference manual BP measurement for each subject, which were used to develop a deep learning-based automatic BP measurement method. Then, the measurement error and BP category classification performance of the deep learning-based method were validated, and its

performance was also compared with the traditional oscillometric method.

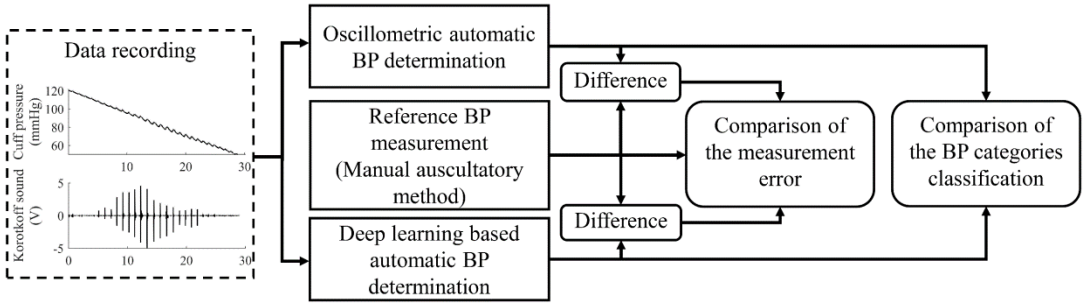

Figure 1. Overall flow diagram of the study methodology**.**

Subjects

Manual BPs were taken from 428 subjects (194 female and 234 male, age of 53 ± 18 years). All subjects gave their written informed consent to participate. This study received ethical permission from the Newcastle & North Tyneside Research Ethics Committee. The investigation conformed with the principles in the Declaration of Helsinki. All the analyses involved were performed on anonymized data.

The subjects were divided randomly into two sub-groups. One sub-group used for developing deep learning-based BP determination algorithm included 314 subjects (147 female and 167 male), aged from 16 to 84 years (56 ± 18 years), while, another sub-group used for validating the performance of the deep learning-based algorithm included 114 subjects (47 female and 67 male), aged from 17 to 84 years (51 ± 18 years). According to the American College of Cardiology (ACC)/American Heart Association (AHA) guidelines for the prevention, detection, evaluation, and management of high blood pressure in adults, subjects were classified into three BP categories: normal (SBP < 120 and DBP < 80 mmHg), elevated (SBP 120 – 129 and DBP < 80 mmHg) and

hypertension (SBP ≥ 130 or DBP ≥ 80 mmHg) categories.[19] The detailed subject demographic information of two sub-groups, including age, gender, height, weight, arm circumference, BP, body mass index and BP distribution are summarized in Table 1.

Table 1. Baseline characteristics of the study subjects

Characteristics	Development Group	Validation Group
No. of total subject	314	114
Female	147 (46.8)	47 (41.2)
Male	167 (53.2)	67 (58.8)
Age (years)	56 ± 18	52 ± 18
Biomarkers		
Height (cm)	168.1 ± 9.4	169.9 ± 11.1
Weight (kg)	73.9 ± 13.2	77.5 ± 18.3
Arm circ (cm)	28.1 ± 2.6	28.7 ± 3.4
SBP (mmHg)	134.7 ± 26.5	129.9 ± 24.8
DBP (mmHg)	73.4 ± 12.3	77.6 ± 15.2
Body mass index (kg/m2)	26.2 ± 4.4	26.7 ± 5.3
BP category		
Normal	89 (28.3)	30 (26.3)
Elevated	91 (29.0)	32 (28.1)
Hypertension	134 (42.7)	52 (45.6)

Values are n (%) or mean ± SD

Reference BP measurement

The manual auscultatory method was employed as a reference for BP measurement in this study. As shown in Figure 2, two trained operators took these measurements simultaneously using a clinically validated manual electronic

sphygmomanometer (Accoson Greenlight 300; AC Cossor & Son (Surgical) Ltd, Harlow, UK). Manual auscultatory SBP and DBP were determined from the appearance and disappearance of the KorS, respectively. The guidelines from the British Hypertension Society and American Heart Association were followed.[6, 20]

BP measurements from each of the two trained operators were reviewed. If the observers' BP reading disagreed with > 4 mmHg in SBP or DBP, another BP measurement was taken. According to the statistical analysis, there was no significant BP difference (for both SBP and DBP) between the BP readings from two observers (both $P > 0.05$). The mean BP values obtained from reference measurements by the observers were then used as the reference BPs for that subject in the following analysis.

For each subject, two valid repeated BP measurements were taken, with a one-minute interval between them, allowing recovery of cardiovascular hemodynamics. All BP measurements were performed in a quiet and temperature-controlled clinical measurement room. Prior to the measurement, each subject had been asked to rest on a chair for 5 minutes and breathe gently during the whole measurement.

Data recording

As shown in Figure 2, during reference manual BP measurement, an automatic and programmable air pump was used to firstly inflate the cuff, and then deflate linearly at the recommended rate of 2-3 mmHg/s. The cuff pressure was recorded by a pressure sensor connected to the cuff via a tube, and the KorS were simultaneously recorded by a bespoke system that included a stethoscope end and a microphone. These analogue

signals were converted to digital signals with a sampling rate of 2000 Hz and a resolution of 16 bits. The final digital KorS and cuff pressure signals were stored in a computer for off-line analysis and processing.

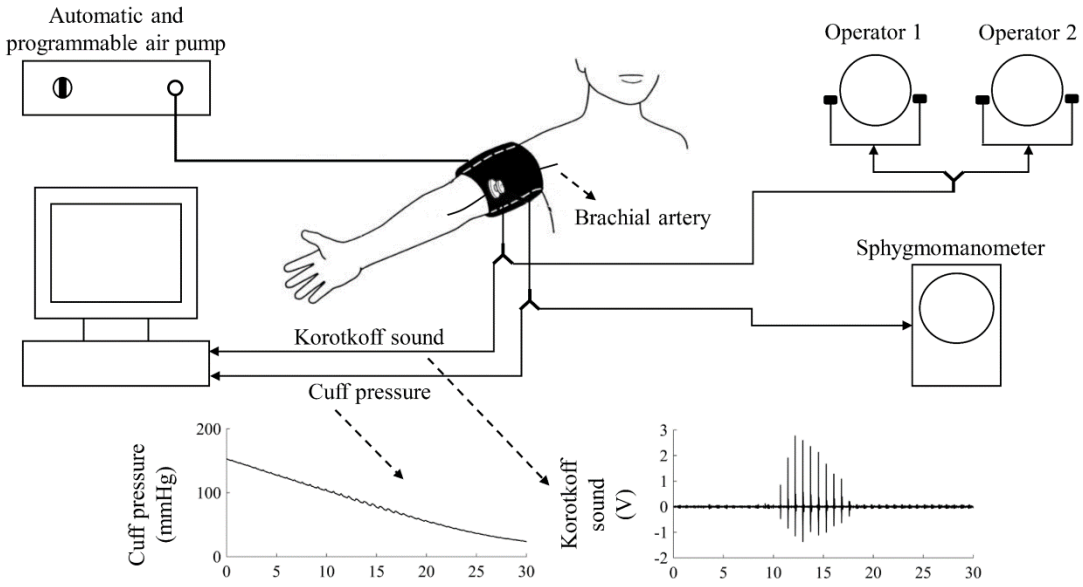

Figure 2. Illustration of the measurement system for simultaneously recording Korotkoff sound and cuff pressure signals. Two trained operators performed the manual auscultatory blood pressure measurement simultaneously. During the measurement the cuff pressure and Korotkoff sounds were digitally recorded and stored in a computer.

Oscillometric automatic BP determination

Oscillometric characteristic ratio determination from development sub-group

Firstly, the oscillometric pulses were extracted from the recorded cuff pressure after segmenting each pulse and removing the baseline cuff pressure. The oscillometric envelope was obtained by fitting a sixth-order polynomial model to the extracted oscillometric pulse peak amplitude, which was plotted against the baseline cuff pressure.

Then, the reference SBP and DBP were used to determine two arterial pulse where the baseline cuff pressure was equal to the reference BPs. Next, the systolic and diastolic characteristic ratios were calculated by dividing the peak amplitude of two arterial pulses by the maximum amplitude of the oscillometric envelope. Finally, the mean systolic and diastolic characteristic ratios were obtained across all subjects in the development sub-group, which were 0.51 and 0.79 for SBP and DBP, respectively. The two ratios were then used for the following oscillometric BP measurement in the validation sub-group.

Oscillometric BP measurement on validation sub-group

The details of oscillometric BP determination have been described in our previously published study.[21] Briefly, as described above, after obtaining the oscillometric envelope, automatic oscillometric MAP was determined from the cuff pressure at the maximum amplitude of the oscillometric envelope. Automatic oscillometric SBP and DBP were determined when the amplitude of oscillometric envelope reached a specific ratio (0.51 for SBP and 0.79 for DBP) of the maximum amplitude of the oscillometric envelop.

Deep learning-based automatic BP determination

Algorithm development to identify audible KorS using development sub-group

There are four steps of our automatic algorithm to identify audible KorS. The first three steps were the pre-processing of KorS signal, to be specific, as shown in figure

3(a), the KorS signal was firstly segmented into beat-by-beat sequenced frames (1s window with 2,000 sample points per frame) according to the peak of oscillometric pulse associated with each cardiac heartbeat. Each frame was then converted into a matrix 'image' (where the x-axis and y-axis represent the time and frequency, respectively, while the value of every pixel indicates the power at a particular time and frequency) by short time Fourier transformation (STFT) with 60 ms Hamming window (sampling rate = 2000 Hz) and 87% overlap. Thirdly, all the frames between the manually determined SBPs and DBPs were labeled as audible KorS beats, while the others were labeled as non-audible KorS beats.

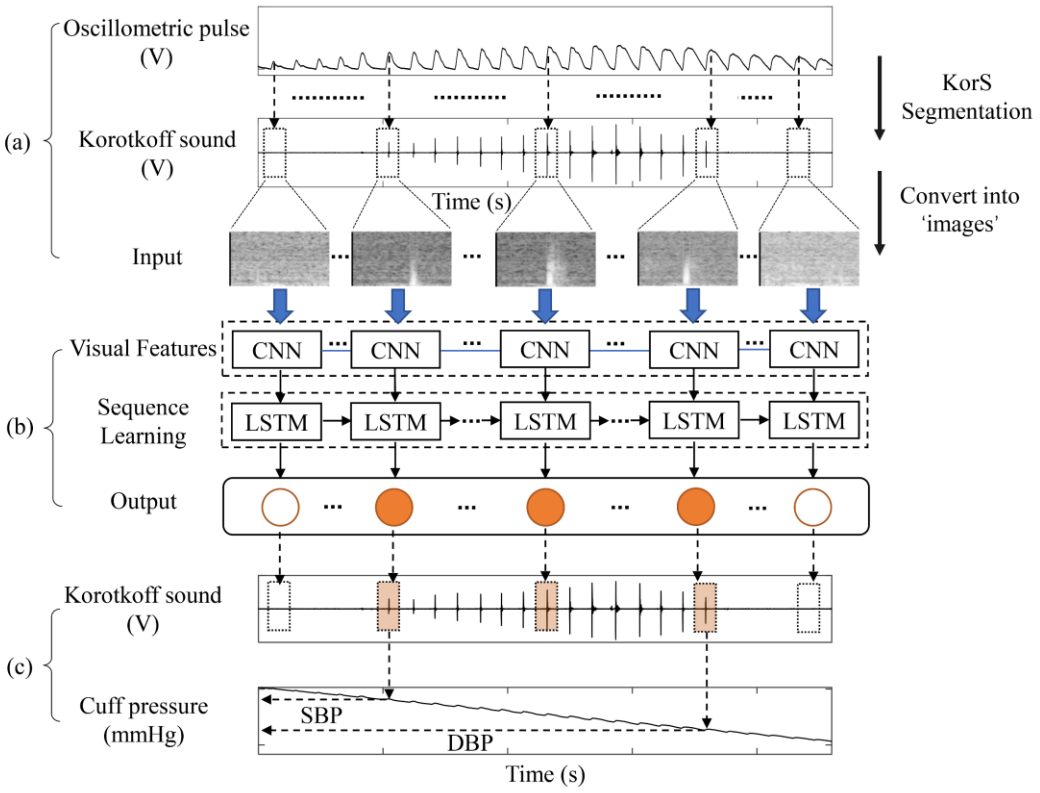

Figure 3. Framework of deep learning-based automatic BP determination algorithm. (a) The pre-processing of KorS signal and the converted KorS time-frequency images as the input of the neural network. (b) The CNN modules learn the features from the

images in sequence. The extracted features were then fed into a LSTM network for temporal features extraction and binary classification. The blue line between CNN modules indicates that these CNN modules share weights. (c) The identified audible KorS beats were used to determine SBP and DBP corresponding to the cuff pressure. (d) The structure of the developed CNN.

During the fourth step (Figure 3(b)), convolutional neural networks (CNNs) were applied to extract features from all frames in sequence and generate a feature sequence for a KorS signal record. Detailly, the CNNs follows standard feature extraction strategies for image-based deep learning tasks such as object recognition, where convolutions layers are used to capture local-to-global features via convoluting information in a sliding window. After every convolution, max-pooling is used to extract the most prominent signal and halve the size of the input, from 76x34 finally to 10x9. After convolution and max-pooling, the features are pulled through three fully-connected layers. The kernel size and stride for the sliding window of each convolution layer are given in Table 2. We in total used three convolution layers and three max-pooling layers. We empirically chose the specific parameters such as kernel size, layer number, etc. to achieve the best results. Then, the feature sequence was fed into a long short-term memory (LSTM) network for binary classification. LSTM is specialized for processing sequential inputs, which can learn the temporal and contextual information of the frames in sequence, especially the dynamics of the input sequence, i.e. the dependencies between consecutive frames.[22] A standard one layer LSTM model was

used to the learn the dynamics. It takes as input the output of the CNNs, then outputs to a fully-connected layer which finally predicts the class label which could perform automatic identification of audible and non-audible KorS beats. The overall structure and setting of the neural networks are given in Table 2.

Table 2. The structure of the proposed neural networks

Layer	*Setting*	*Output*
Convolution	8@3×3/1,1	64×76×34×8
Max-Pooling	2×2/2,2	64×38×17×8
Convolution	16@3×3/1,1	64×38×17×16
Max-Pooling	2×2/2,2	64×19×9×16
Convolution	16@3×3/1,1	64×19×9×16
Max-Pooling	2×2/2,1	64×10×9×16
Fully Connected	-	64×1440
Fully Connected	-	64×96
Fully Connected	-	64×96
LSTM		64×96
Fully Connected	-	64×2

The setting is presented as: number of filters @ kernel size / stride

During the training process, the parameters of the neural networks were initialized random values. Then, for each input signal, the prediction given by the neural networks

was compared with the known label from development sub-group, and parameters of the neural networks were then modified to decrease the error on that signal (adaptive moment estimation). This process was repeated for every signal record in the development sub-group until the neural networks 'learn' how to accurately identify audible and non-audible KorS. The training process were performed on a computer with CPU (AMD Ryzen 5 2600 @ 3.4 GHz) and GPU (NVIDIA GTX 1080).

Deep learning-based BP measurement on the validation sub-group

The overall framework of our deep learning-based automatic BP determination algorithm is shown in Figure 3. The trained neural networks with fixed parameters were used to identify the audible and non-audible KorS beats. Then, a mapping algorithm was developed to associate the identified Korotkoff beats for BP determination (Figure 3(c)). The first and last identified audible KorS beats were used, respectively, to determine SBP and DBP corresponding to the baseline cuff pressure.[17] In order to follow the guideline of manual auscultatory BP measure, an additional determination rule was applied: SBP was determined with at least two consecutive identified audible KorS beats, and DBP was determined at the point at which all sounds finally disappear completely.

Data and statistical analysis

For each subject, there were two repeated measurements and three BP determinations (1 from the reference manual BP measurement, 1 from oscillometric

method and 1 from deep learning method). In total, there were 684 SBP and 684 DBP values from 114 validation subjects (684 = 114 subject * 2 repeats * 3 determination methods).

The SPSS Statics 19 software package (SPSS Inc, Chicago, IL, USA) was employed in this study to perform the statistical analysis and a *P* value less than 0.05 was considered to be significant.

The fusion matrix was firstly employed to evaluate the performance of proposed algorithm to identify audible and non-audible KorS beats. Then, The BPs obtained by the oscillometric method were compared to the reference measurement to calculate their mean paired BP difference (measurement error) and standard deviation (SD) of their differences, respectively for all the validation subjects as well as for the three BP categories (normal, elevated and hypertension). The same data analysis approach was performed for the deep learning method. Analysis of variance analysis (ANOVA) method was used to investigate the repeatability between the 2 repeated measurements, and the effect of the three measurement methods on the obtained BPs with post-hoc paired comparison. Next, the histograms of BP differences across all the validation subjects, the distribution of absolute BP differences within 5, 10 and 15 mmHg and the Bland-Altman scatterplots were obtained.

Finally, the performance of the oscillometric and deep learning methods were compared in terms of their ability to correctly classify the subject into three categories (normal, elevated and hypertension). Three standard metrics (classification sensitivity, specificity and accuracy) were calculated from the calculation of true positive (TP),

true negative (TP), false positive (FP) and false negative (FN), where the sensitivity is the ratio of correctly classified events (i.e., normal, elevated or hypertensive subjects) among all the events, $Sensitivity = TP / (TP + FN)$; the specificity is the radio of correctly classified nonevents (i.e., non-normal, non-elevated or non-hypertensive subjects) among all of the nonevents, $Specificity = TN / (TN + FP)$; and the accuracy is the ratio of the number of correctly classified subjects to the total number of subjects classified, $Accuracy = (TP + TN) / (TP + TN + FP + FN)$.

Results

Performance of the identification of audible and non-audible KorS beat

The confusion matrix on validation sub-group obtained by the proposed neural networks has been given in Table 3, where 253 audible KorS beats were misidentified as non-audible KorS beats and 462 non-audible KorS beats were misidentified as audible KorS beats, and the identification accuracy is 95.1%.

Table 3. Confusion matrix for the KorS beat identification of the validation sub-group

	Audible KorS beats	Non-audible KorS beats
Audible KorS beats	4511	253
Non-audible KorS beats	462	9366

Comparison of the measurement error from the oscillometric method and the deep learning method

According to the statistical analysis, there was no significant BP difference (SBP and DBP) between the oscillometric method and the reference manual BP measurement (all $P > 0.3$). Similarity, there was no significant BP difference determined between the deep learning method and the reference manual BP measurement (all $P > 0.5$).

The automatic BP differences (mean ± SD) in comparison with the reference manual BPs are given in Table 4, separately for the oscillometric and deep learning methods. This is respectively given for the three BP categories. The overall BP differences of oscillometric method were -0.4 mmHg and 0.2 mmHg, respectively, for SBP and DBP, while the deep learning method achieved 0.2 mmHg and 0.1 mmHg, respectively, for SBP and DBP. More interestingly, the deep learning method produced smaller SD of difference (with the range of 2.6 to 4.8 mmHg for SBP and DBP from all three BP categories) than those of oscillometric method (with the range of 4.4 to 6.7 mmHg).

Table 4. The mean ± SD of BP differences for three BP categories determined by the oscillometric method and deep learning method in reference to the reference manual BP measurement. The overall mean ± SD of BP differences across all subjects are also presented.

	BP difference between Oscillometric and Reference (mmHg)		*BP difference between Deep learning and Reference (mmHg)*	
	SBP	DBP	SBP	DBP
Normal	2.1 ± 6.0	-0.6 ± 4.4	-0.2 ± 4.4	0.8 ± 3.6

Elevated	-2.2 ± 5.5	-0.5 ± 5.7	-1.0 ± 4.2	-0.5 ± 2.6
Hypertension	-0.7 ± 6.7	1.0 ± 6.2	1.1 ± 4.8	0.2 ± 3.2
Overall	-0.4 ± 6.4	0.2 ± 5.7	**0.2 ± 4.6**	**0.1 ± 3.2**

Figure 4 given the Bland-Altman scatterplots of the SBP and DBP determined by the automatic oscillometric method and deep learning method versus the reference manual BP values. The limits of agreement for the deep learning method (-8.8 to 9.2 mmHg for SBP (Figure 4b) and -6.2 to 6.4 mmHg for DBP (Figure 4d)) were smaller than those corresponding values from the oscillometric method (-12.9 to 12.1 mmHg for SBP (Figure 4a) and -11.0 to 11.4 mmHg for DBP (Figure 4c)).

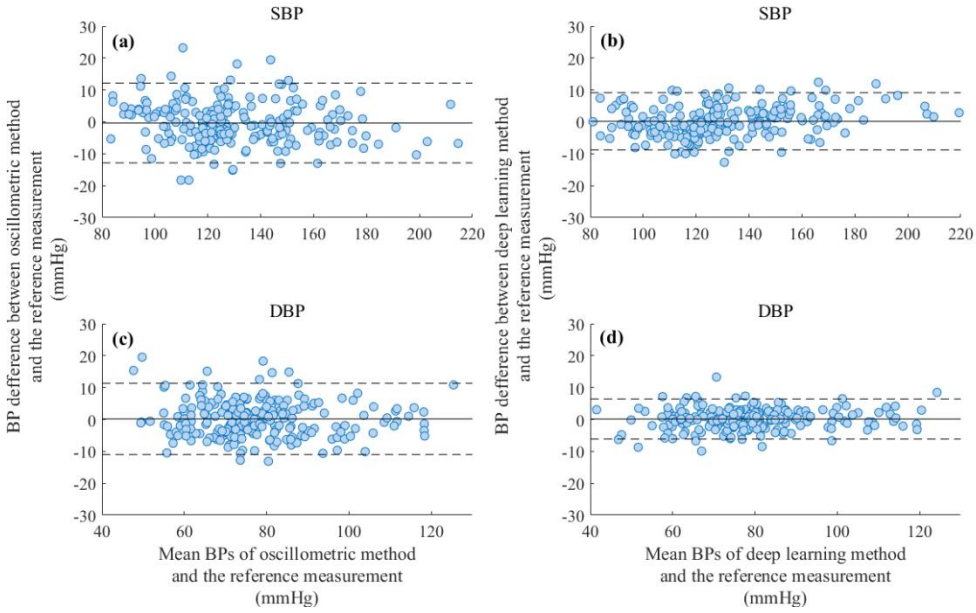

Figure 4. Bland-Altman plots of (a) SBP and (c) DBP from the oscillometric method versus reference manual BP measurement, and (b) SBP and (d) DBP from the deep

learning method versus reference manual BP measurement. The limits of agreement (1.96 * SD of BP difference) are given using the dashed lines in figures.

The proportion of BP differences within 5, 10 and 15 mmHg are shown in Figure 5 and Table 5. It can be observed that, for each level (i.e. within 5, 10 and 15 mmHg), the proportion of BP differences (both SBP and DBP) obtained by the deep learning method were within the Grade A standard for BP device by BHS (60%, 85% and 95% of BP differences are within 5, 10 and 15 mmHg, respectively).[23] Although the oscillometric method also achieved a grade of A for DBP, the results for SBP were in a grade of B (50%, 75% and 90% of BP differences are within 5, 10 and 15 mmHg, respectively).

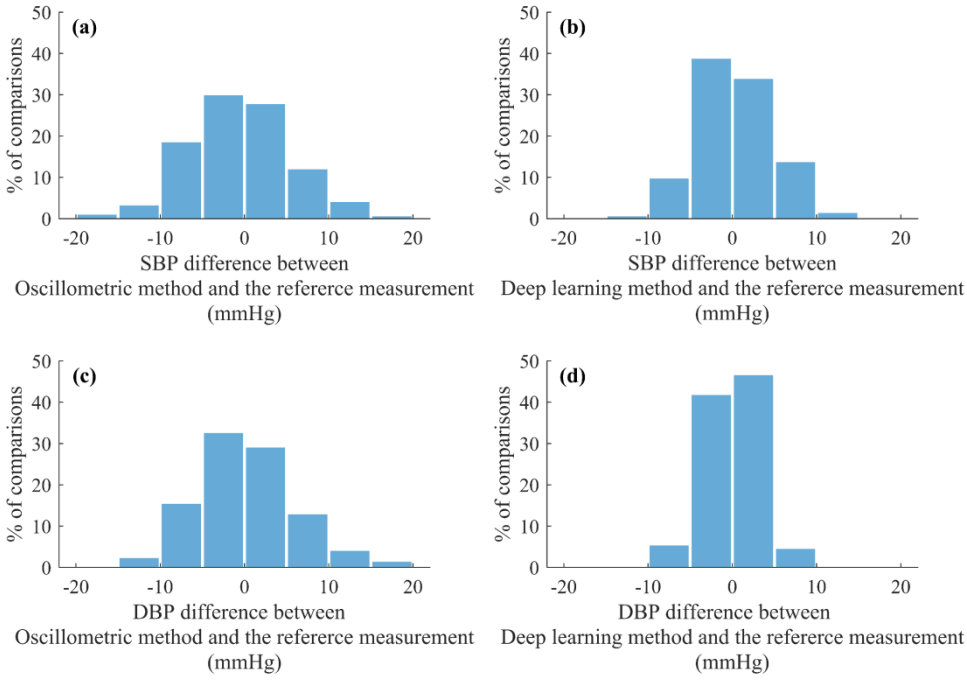

Figure 5. Histograms of within-subject (a) SBP and (c) DBP differences between the oscillometric method and the reference manual BP measurement. And histograms of within-subject (b) SBP and (d) DBP differences between the deep learning method and the reference manual BP measurement.

Table 5. Distributions of BP differences between the oscillometric method and the reference manual BP measurement, and between the deep learning method and the reference manual BP measurement.

	Oscillometric method		*Deep learning method*	
	SBP (%)	DBP (%)	SBP (%)	DBP (%)
Within 5 mmHg	58.3	62.3	**73.2**	**89.0**
Within 10 mmHg	89.5	91.2	**97.4**	**99.6**
Within 15 mmHg	97.4	98.2	**100.0**	**100.0**

Evaluation results for classifying BP categories

As shown in Table 6, against the reference measurement (manual auscultatory method, the gold standard of noninvasive BP measurement), both automatic oscillometric and deep learning methods had the same performance to identify the normal and elevated categories. More importantly, the deep learning method achieved higher sensitivity, specificity and accuracy for classifying hypertensive categories (all metrics higher than 90%) than the oscillometric method.

Table 6. Classification of BP categories by the oscillometric method and the deep learning method along with sensitivity, specificity and accuracy (against the reference measurement, gold standard of noninvasive BP measurement).

BP Category	*TP*	*FN*	*TN*	*FP*	*Sensitivity (%)*	*Specificity (%)*	*Accuracy (%)*
Oscillometric method							
Normal	30	0	74	10	100.0	88.1	91.2
Elevated	20	12	82	0	62.5	100.0	89.5
Hypertension	42	10	59	3	**80.8**	**95.2**	**88.6**
Deep learning method							
Normal	30	0	74	10	100.0	88.1	91.2
Elevated	20	12	82	0	62.5	100.0	89.5
Hypertension	47	5	60	2	**90.4**	**96.8**	**93.9**

Abbreviations: TP, true positive; FN, false negative; TN, true negative; FP, false positive. The BP category defined as: Normal (SBP < 120 and DBP < 80 mmHg), Elevated (SBP 120 – 129 and DBP < 80 mmHg) and Hypertension (SBP ≥ 130 or DBP ≥ 80 mmHg).

Discussions

This study clinically validated the performance of a deep learning-based method on subjects with a wide range of BPs. Its performance was compared with the manual reference method and the widely used automatic oscillometric method. Across all the subjects, the deep learning method achieved overall BP measurement errors of 0.2 mmHg for SBP and 0.1 mmHg for DBP, where the SD of BP difference from deep learning method were smaller than those of the oscillometric method (4.6 vs 6.4 mmHg

for SBP and 3.2 vs 5.7 mmHg for DBP). This was also demonstrated from the Bland-Altman plots, where the smaller limits of agreement were obtained from the deep learning method. The better performance could be explained by the different measurement principles used for the deep learning-based auscultatory method and oscillometric method. The oscillometric method is based on empirical equations to estimate BPs. Our proposed automatic deep learning auscultatory method automates the principle of the gold standard of non-invasive BP measurement.[24] In theory, the outcome of our proposed method could identical to the manual auscultatory method if the performance of our developed neural networks on identifying the audible and non-audible KorS beats was perfect. Our previously published has evaluated the potential variation of using CNN to identify audible KorS during BP measurement.[25] Nevertheless, better measurement performance with smaller SD of BP difference has been achieved by the deep learning method than the oscillometric method.

It was observed that, in comparison with the oscillometric method, the deep learning method achieved better performance for classifying hypertension category. This suggests that, the deep learning method had higher ability to correctly classify a patient with hypertension as a hypertensive individual, whereas to correctly classify a patient with no hypertension as a non-hypertensive individual. An increase in arterial stiffness with reductions in arterial compliance, known to occur in older patients, might have contributed to the explanation of the discrepancy.[8, 26] In hypertensive case, KorS are still audible whereas the oscillations could be reduced due to artery stiffening,[11] lading to potential measurement variability.

It is worth noting that, both oscillometric and deep learning methods had low sensitivity (both were 62.5 %) for classifying elevated subjects, which indicates that their ability of correctly identifying an elevated patient is not strong enough. However, the deep learning method achieved smaller measurement error (-1.0 vs -2.2 mmHg for SBP) and smaller SD of BP differences (4.2 vs 5.5 mmHg for SBP and 2.6 vs 5.7 mmHg for DBP) on elevated subjects than those of oscillometric method. The misclassification was observed with subjects whose determined BPs were close to the edge of the BP range of the elevated category. For example, if SBP of a subject is 120 mmHg or 122 mmHg, even a little underestimated could result in classifying it incorrectly as normotensive. This indicates that, repeated measurement is a very important strategy for the classification of the BP categories, especially, in the case that the determined BP is close to an edge of the BP ranges.

One limitation of this study is that the effect on measurement accuracy of cardiovascular disease or comorbidities (such as obesity, diabetes, hyperlipidemia, or peripheral vascular disease) has not been analyzed comprehensively. Future studies on focused clinical groups with cardiovascular diseases and comorbidities are suggested to investigate whether the measurement error would increase in these diseases in a different manner.

Another limitation is that, non-invasive BP measurement is perfect in comparison with the true invasive reference measurement. Picone D et al has reported that oscillometric BP and auscultatory BP systematically underestimated intra-arterial brachial SBP, and overestimated intra-arterial brachial DBP.[27] However, the manual

auscultatory method is regarded as gold standard of non-invasive BP measurement. It has been widely accepted and used for automatic BP device validation. A future study with an invasive intra-arterial reference standard could be proposed.

Conclusion

In summary, this study has provided clinical evidence that our proposed deep learning-based automatic auscultatory method can achieve accurate measurement and high BP category classification performance, demonstrating that it can be developed further to replace the automatic oscillometric and manual auscultatory method.

Acknowledgments

The authors acknowledge the supporting of Newcastle University Department of regional medical physics and cardiovascular physics engineering research group's staff for supporting the study.

Funding

This work was supported by China Postdoctoral Science Foundation (Grant number 2019M653409), Chengdu Science and Technology Bureau (Grant number 2019-YF05-00109-SN), Sichuan Science and Technology Program (Grant number 2020YJ0282). The experiment was conducted with the support from the Engineering and Physical Sciences Research Council (EPSRC) Healthcare Partnership Award (Grant number EP/I027270/1).

Conflict of interest statement

The Authors declare that there is no conflict of interest.

References

1. Collaborators GBDRF. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017;390:1345-1422.
2. Lawes CM, Vander Hoorn S, Law MR, Elliott P, MacMahon S and Rodgers A. Blood pressure and the global burden of disease 2000. Part II: estimates of attributable burden. *J Hypertens*. 2006;24:423-430.
3. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, Clement DL, Coca A, de Simone G, Dominiczak A, Kahan T, Mahfoud F, Redon J, Ruilope L, Zanchetti A, Kerins M, Kjeldsen SE, Kreutz R, Laurent S, Lip GYH, McManus R, Narkiewicz K, Ruschitzka F, Schmieder RE, Shlyakhto E, Tsioufis C, Aboyans V, Desormais I and Authors/Task Force M. 2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension. *J Hypertens*. 2018;36:1953-2041.
4. Lenfant C, Chobanian AV, Jones DW, Roccella EJ, Joint National Committee on the Prevention DE and Treatment of High Blood P. Seventh report of the Joint National Committee on the Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7): resetting the hypertension sails. *Hypertension*.

2003;41:1178-1179.

5. Jones DW, Appel LJ, Sheps SG, Roccella EJ and Lenfant C. Measuring blood pressure accurately: new and persistent challenges. *JAMA*. 2003;289:1027-1030.
6. Beevers G, Lip GY and O'Brien E. ABC of hypertension: Blood pressure measurement. Part II-conventional sphygmomanometry: technique of auscultatory blood pressure measurement. *BMJ*. 2001;322:1043-1047.
7. O'Brien E and Fitzgerald D. The history of blood pressure measurement. *J Hum Hypertens*. 1994;8:73-84.
8. Muntner P, Shimbo D, Carey RM, Charleston JB, Gaillard T, Misra S, Myers MG, Ogedegbe G, Schwartz JE, Townsend RR, Urbina EM, Viera AJ, White WB and Wright JT, Jr. Measurement of Blood Pressure in Humans: A Scientific Statement From the American Heart Association. *Hypertension*. 2019;73:e35-e66.
9. Alpert BS, Quinn D and Gallick D. Oscillometric blood pressure: a review for clinicians. *J Am Soc Hypertens*. 2014;8:930-938.
10. Popele NMv, Bos WJW, Beer NAMd, Kuip DAMvd, Hofman A, Grobbee DE and Witteman JCM. Arterial Stiffness as Underlying Mechanism of Disagreement Between an Oscillometric Blood Pressure Monitor and a Sphygmomanometer. *Hypertension*. 2000;36:484-488.
11. van Montfrans GA. Oscillometric blood pressure measurement: progress and problems. *Blood Pressure Monitoring*. 2001;6:287-290.
12. Jones DW and Hall JE. Hypertension: Pathway to Success. *Hypertension*. 2008;51:1249-1251.

13. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL and Webster DR. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016;316:2402-2410.

14. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP and Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. 2019;25:65-69.

15. Ko WY, Siontis KC, Attia ZI, Carter RE, Kapa S, Ommen SR, Demuth SJ, Ackerman MJ, Gersh BJ, Arruda-Olson AM, Geske JB, Asirvatham SJ, Lopez-Jimenez F, Nishimura RA, Friedman PA and Noseworthy PA. Detection of Hypertrophic Cardiomyopathy Using a Convolutional Neural Network-Enabled Electrocardiogram. *J Am Coll Cardiol*. 2020;75:722-733.

16. Parisot S, Ktena SI, Ferrante E, Lee M, Guerrero R, Glocker B and Rueckert D. Disease prediction using graph convolutional networks: Application to Autism Spectrum Disorder and Alzheimer's disease. *Medical Image Analysis*. 2018;48:117-130.

17. Pan F, He P, Chen F, Zhang J, Wang H and Zheng D. A novel deep learning based automatic auscultatory method to measure blood pressure. *Int J Med Inform*. 2019;128:71-78.

18. Pan F, He P, Chen F, Pu X, Zhao Q and Zheng D. Deep learning-based automatic blood pressure measurement: evaluation of the effect of deep breathing, talking

and arm movement. *Ann Med*. 2019;51:397-403.

19. Whelton PK, Carey RM, Aronow WS, Casey DE, Jr., Collins KJ, Dennison Himmelfarb C, DePalma SM, Gidding S, Jamerson KA, Jones DW, MacLaughlin EJ, Muntner P, Ovbiagele B, Smith SC, Jr., Spencer CC, Stafford RS, Taler SJ, Thomas RJ, Williams KA, Sr., Williamson JD and Wright JT, Jr. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol*. 2018;71:2199-2269.
20. Pickering TG, Hall JE, Appel LJ, Falkner BE, Graves J, Hill MN, Jones DW, Kurtz T, Sheps SG and Roccella EJ. 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. *Circulation*. 2005;111:697-716.
21. Zheng D, Pan F and Murray A. Effect of mechanical behaviour of the brachial artery on blood pressure measurement during both cuff inflation and cuff deflation. *Blood Press Monit*. 2013;18:265-271.
22. Hochreiter S and Schmidhuber J. Long Short-Term Memory. *Neural Computation*. 1997;9:1735-1780.
23. O'Brien E, Waeber B, Parati G, Staessen J and Myers MG. Blood pressure

measuring devices: recommendations of the European Society of Hypertension. *BMJ*. 2001;322:531-536.

24. Stergiou GS, Alpert B, Mieke S, Asmar R, Atkins N, Eckert S, Frick G, Friedman B, Grassl T, Ichikawa T, Ioannidis JP, Lacy P, McManus R, Murray A, Myers M, Palatini P, Parati G, Quinn D, Sarkis J, Shennan A, Usuda T, Wang J, Wu CO and O'Brien E. A Universal Standard for the Validation of Blood Pressure Measuring Devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) Collaboration Statement. *Hypertension*. 2018;71:368-374.
25. Pan F, He P, Liu C, Li T, Murray A and Zheng D. Variation of the Korotkoff Stethoscope Sounds During Blood Pressure Measurement: Analysis Using a Convolutional Neural Network. *IEEE J Biomed Health Inform*. 2017;21:1593-1598.
26. Cheitlin MD. Cardiovascular physiology-changes with aging. *Am J Geriatr Cardiol*. 2003;12:9-13.
27. Picone DS, Schultz MG, Otahal P, Aakhus S, Al-Jumaily AM, Black JA, Bos WJ, Chambers JB, Chen CH, Cheng HM, Cremer A, Davies JE, Dwyer N, Gould BA, Hughes AD, Lacy PS, Laugesen E, Liang F, Melamed R, Muecke S, Ohte N, Okada S, Omboni S, Ott C, Peng X, Pereira T, Pucci G, Rajani R, Roberts-Thomson P, Rossen NB, Sueta D, Sinha MD, Schmieder RE, Smulyan H, Srikanth VK, Stewart R, Stouffer GA, Takazawa K, Wang J, Westerhof BE, Weber F, Weber T, Williams B, Yamada H, Yamamoto E and Sharman JE. Accuracy of Cuff-

Measured Blood Pressure: Systematic Reviews and Meta-Analyses. *J Am Coll Cardiol*. 2017;70:572-586.