Computer Methods and Programs in Biomedicine 207 (2021) 106191

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

A hybrid neural network for continuous and non-invasive estimation of blood pressure from raw electrocardiogram and photoplethysmogram waveforms

Stephanie Baker[a,*], Wei Xiang[b], Ian Atkinson[c]

[a] College of Science & Engineering, James Cook University, Cairns, Queensland, Australia 4878, Australia
[b] School of Engineering and Mathematical Sciences, La Trobe University, Melbourne, Victoria, Australia 3086, Australia
[c] eResearch Centre, James Cook University, Townsville, Queensland, Australia 4811, Australia

ARTICLE INFO

Article history:
Received 25 January 2021
Accepted 12 May 2021

Keywords:
Cuffless blood pressure
Photoplethysmogram
Electrocardiogram
Machine learning
Neural networks
Wearable healthcare

ABSTRACT

Background and objectives: Continuous and non-invasive blood pressure monitoring would revolutionize healthcare. Currently, blood pressure (BP) can only be accurately monitored using obtrusive cuff-based devices or invasive intra-arterial monitoring. In this work, we propose a novel hybrid neural network for the accurate estimation of blood pressure (BP) using only non-invasive electrocardiogram (ECG) and photoplethysmogram (PPG) waveforms as inputs.

Methods: This work proposes a hybrid neural network combines the feature detection abilities of temporal convolutional layers with the strong performance on sequential data offered by long short-term memory layers. Raw electrocardiogram and photoplethysmogram waveforms are concatenated and used as network inputs. The network was developed using the TensorFlow framework. Our scheme is analysed and compared to the literature in terms of well known standards set by the British Hypertension Society (BHS) and the Association for the Advancement of Medical Instrumentation (AAMI).

Results: Our scheme achieves extremely low mean absolute errors (MAEs) of 4.41 mmHg for SBP, 2.91 mmHg for DBP, and 2.77 mmHg for MAP. A strong level of agreement between our scheme and the gold-standard intra-arterial monitoring is shown through Bland Altman and regression plots. Additionally, the standard for BP devices established by AAMI is met by our scheme. We also achieve a grade of 'A' based on the criteria outlined by the BHS protocol for BP devices.

Conclusions: Our CNN-LSTM network outperforms current state-of-the-art schemes for non-invasive BP measurement from PPG and ECG waveforms. These results provide an effective machine learning approach that could readily be implemented into non-invasive wearable devices for use in continuous clinical and at-home monitoring.

1. Introduction

Blood pressure (BP) is a key diagnostic tool for a variety of life-threatening conditions. Elevated BP, or hypertension, is a major risk factor for cardiovascular disease (CVD), contributing to the deaths of 9.4 million people every year [1]. Additionally, poor organ perfusion can be identified through the measurement of BP-derived parameters, particularly mean arterial pressure (MAP). MAP is useful in determining overall blood flow and thus the level of nutrient delivery to organs, and thus is routinely measured when dealing with high-mortality conditions like septic shock [2]. MAP that is too low can lead to shock, syncope, and poor perfusion to organs, while elevated MAP places strain on the cardiovascular system and can eventually to various CVDs including stroke [3].

Despite the importance of monitoring BP, there are currently no commercially available devices capable of continuous and non-invasive BP measurement that have been approved for medical use. Currently, the gold-standard method for continuous and accurate BP monitoring is intra-arterial monitoring, which involves the invasive insertion of an arterial line into a patient's artery [4]. This is clearly not suitable for long-term monitoring as it must be performed in a clinical environment and it increases infection risk for

* Corresponding author.
E-mail addresses: stephanie.baker@jcu.edu.au (S. Baker), w.xiang@latrobe.edu.au (W. Xiang), ian.atkinson@jcu.edu.au (I. Atkinson).

https://doi.org/10.1016/j.cmpb.2021.106191

the patient. Typically, BP is measured using less invasive sphygmomanometers, cuff-based devices which is manually or automatically inflated to determine BP. However, sphygmomanometers are incapable of continuous monitoring and cause significant discomfort to many patients [5]. They also cannot be used on people with several pre-existing conditions, such as lymphedema [6].

There is a clear need for improved methods of clinical and at-home BP monitoring, especially for high-risk patients. As such, many recent works have investigated methods for non-invasive measurement of BP. One promising area of research lies in machine learning (ML). ML techniques have been used to estimate BP from various health factors such as age and gender [7], as well as for improving sphygmomanometer measurements [8,9].

More recently, many researchers have investigated the calculation of BP from electrocardiogram (ECG) and photoplethysmogram (PPG) signals [10–20]. In [10,11,19], the Medical Information Mart for Intensive Care III (MIMIC III) database was used to obtain features such as pulse transit time (PTT) and other manually extracted features of the ECG and PPG waveforms. These were then used with algorithms including AdaBoost in [10], multi-regression in [11] and multivariate adaptive regression spline (MARS) analysis in [19]. Unfortunately, the MIMIC III database used by these works suffers from intra-waveform alignment issues that make calculation of PTT and other time-dependent features between ECG and PPG signals unreliable [21]. As such, these schemes could not be reliably applied in healthcare applications.

In [12], raw PPG waveforms are used to train an AdaBoostR algorithm to estimate SBP, DBP and MAP. This model was trained on a small subset of the MIMIC-II data of 1,323 records and not validated on a distinct testing set. In results presented from the training set, the model was clearly suffering from a large number of high-range errors and large standard deviation (SD), particularly in SBP estimation. This indicates that the model had overfit to the training data, and therefore would be unlikely to perform strongly on new data.

Meanwhile, in the recent paper [20], raw ECG waveforms are used to train a neural network for SBP, DBP, and MAP prediction. Testing was performed on MIMIC III data, as well as a second independent database. Results across these two databases varied significantly, with the scheme shown to not perform as strongly on the large MIMIC III database. It is likely that utilizing both PPG and ECG data would significantly improve performance, and PPG signals are comparatively easy to obtain from wearable devices compared to ECG signals. Obtaining both signals is highly feasible in hospitals, where both signals are routinely recorded. It is also becoming increasingly feasible in wearable technology with PPG having long been available in smartwatches, and with recent generations of devices such as the Samsung Galaxy Watch [22] and Apple Watch [23] now offering ECG functionality.

Two recent works [13,14] obtained features from ECG and PPG signals before using ML techniques to predict BP. Features considered included PTT, which was measured using the same equipment across all participants. This likely improved synchronization between devices when compared to the waveform synchronity issues in [10,11,24], however clock drift could still impair PTT calculation over longer periods. Each of these works built small databases using measurements from healthy volunteers, with [13] obtaining readings from 85 patients and [14] using 20-second segments from 110 subjects. Good results were presented in both works, however larger databases would be needed to verify that these schemes would perform well on a wide range of patients.

In this paper, we propose a hybrid deep neural network (DNN) that incorporates temporal convolutional neural network (CNN) and long short-term memory (LSTM) layers for the estimation of SBP, DBP, and MAP from raw ECG and PPG waveforms with duration of 5 s. The CNN layers act to identify the important features of the waveform and thus reduce dimensionality, while the LSTM layers are included for their ability to remember information and identify relationships between features to determine a final estimate for SBP and DBP.

Through using raw waveforms as inputs, we allow the hybrid CNN-LSTM network to learn from all of the available information, rather than from manually identified features. This means that our work does not rely on complex time-dependent features such as PTT, thus avoiding issues associated with clock drift between devices. By avoiding manual feature selection, we also introduce the advantages of removing human bias from training and testing. Additionally, the use of short 5-second ECG and PPG waveforms ensures that SBP, DBP, and MAP can be calculated rapidly and continuously, with a lower risk of interference than is seen in longer windows.

The remainder of this paper is structured as follows. Section II presents our methodology, including our schemes for preprocessing, signal quality assessment and developing the hybrid DNN for BP estimation. Section III discusses the results of testing conducted on the DNN algorithms to assess their performance. Finally, Section IV briefly concludes this work and its significance.

2. Methodology

2.1. Data acquisition

Deep learning is most successful when large quantities of data are used for training, validating, and testing the models. The Medical Information Mart for Intensive Care (MIMIC) [25] database features many de-identified patient records from critical care environments and has been used in several significant and high-impact studies to develop biomedical algorithms, including [10,11]. Patient demographics are not available for many waveform records in MIMIC-III, however the overall patient demographics for the entire MIMIC-III database are presented in the original paper describing the data [25].

To train neural networks to estimate SBP, DBP, and MAP, we chose to use only ECG and PPG waveforms as inputs - as such, these waveforms were required. Additionally, reference ground truth values for SBP and DBP are required to give the neural network an output target during training and to allow for assessment of network performance during testing. SBP and DBP are derived from the arterial blood pressure (ABP) waveforms available in the MIMIC database. As such, all records that contained ECG, PPG and ABP waveforms were obtained, resulting in a database comprised of 6,972 unique patients.

2.2. Data preprocessing

Following acquisition, each record was split into 5-second segments. This segment length allows for extremely rapid calculation of SBP, DBP, and MAP, while also providing a wide enough window to accurately calculate the aforementioned BP parameters even where the heart rate is extremely low. While one previous work [15] determined the segment length by the number of peaks and thereafter resampled the segment to fit within a fixed input feature vector length, a temporal approach is favoured by many works [12,14,17]. In this work, a temporal approach to window sizing is taken as this minimises pre-processing and thus improves real-time computational performance. A temporal approach also ensures a consistent sampling frequency is used for all data and preserves time-domain information, improving the network's ability to learn from the data.

Segments were taken sequentially, with no overlap between segments. While using overlapping segments is a viable technique

for data augmentation in smaller databases, it was not necessary for this work given the large size of the MIMIC-III database. Extracting non-overlapping segments ensured that each segment contained completely unique data, reducing the risk of overfitting and minimising the impact of any erroneous data that remained after data preprocessing. During the segmentation process, any segments with missing or flatlining signals were immediately discarded. No further pre-processing of signals was undertaken in this work, however the signals obtained from MIMIC-III were recorded in hospital environment and thus were likely denoised by the high-quality equipment that recorded them.

2.3. Data selection

Signal quality indices (SQIs) have been developed in several previous works to assess the quality of ECG signals, using techniques including spectral analysis [26], fuzzy support vector machines [24,27], and simple sanity checks [28].

While these works offer significant SQI tools for ECG signal assessment, they do not consider PPG signals. As such, for this work we implement a straightforward SQI strategy comprised of sanity checks for PPG and ECG waveforms. Heart rate (HR), beat-to-beat (BTB) intervals and waveform heights are calculated for ECG and PPG in each record. HR values derived from each signal must be equal and fall within 40-180 bpm for a record to be considered "good" by our SQI tool. This is the range which is physiologically probable for HR [28] and thus is a good indicator of signal quality. The consistency of the signal is also considered by finding the maximum-to-minimum ratio for both beat-to-beat intervals and peak heights and ensuring that the maximum is no more than 50% larger than the minimum.

Records were also excluded if pulse pressure (the difference between SBP and DBP) was not between 20–60 mmHg. Pulse pressure is considered high when it is over 60 mmHg [29,30], and is usually indicative of an immediate health problem. Meanwhile, pulse pressure is considered low beneath 40 mmHg and indicates poor heart function [29,30], so an intentionally conservative lower limit of 20 mmHg was chosen for this application given that data is acquired from critical care units. While records containing physiologically improbable pulse pressures are excluded from the training database to ensure data quality, the neural network is not dependent on this feature or any other ABP features, and thus would likely adapt successfully if it encounters valid ECG and PPG data from a patient with exceptionally high or low pulse pressure.

In Algorithm 1, *hr_ppg* and *hr_ecg* are the HRs calculated from the PPG and ECG signal respectively. Additionally, *ppg_peak_ratio* and *ecg_peak_ratio* are the ratios of the maximum peak height to the minimum peak height for each signal, while *ppg_btb_ratio* and *ecg_btb_ratio* are the ratios of the widest to smallest BTB intervals for each signal. Each of these metrics offers a measure of signal consistency, which in turn is indicative of signal quality. Lastly, the parameter *pulse_pressure* is the pulse pressure values calculated from the ABP signal.

Algorithm 1 Signal Selection Algorithm.

Input: hr_ppg, hr_ecg, ppg_peak_ratio, ecg_peak_ratio, ppg_btb_ratio, ecg_btb_ratio, true_sbp, true_dbp, pulse_pressure
Output: use_record
1: **if** (hr_ppg == hr_ecg) & (hr_ppg > 40) & (hr_ppg < 180) & (ppg_peak_ratio < 1.5) & (ecg_peak_ratio < 1.5) & (ppg_btb_ratio < 1.5) & (ecg_btb_ratio < 1.5) & (pulse_pressure > 20 & (pulse_pressure < 60)] **then**
2: record_quality = 1
3: **else**
4: record_quality = 0
5: **end if**

After assessing the suitability of all signals using Algorithm 1, the resulting data was inspected and outlier BP values were excluded. The final database contained over 200,000 records for use in training and testing of the proposed DNNs.

For each "good" record, a feature vector was developed by unravelling both the PPG and ECG signals into vectors of their respective amplitudes. These vectors were then joined to create a single feature vector that contained all amplitude information for the PPG and ECG signals consecutively. The input feature vector thus contained only raw ECG and PPG signals.

2.4. Proposed neural network

For the task of blood pressure estimation from raw waveforms, we propose a hybridised deep neural network that combines temporal convolutional layers with long short-term memory (LSTM) layers, as shown in Fig. 1. CNNs are typically used to identify important features and patterns within a signal, and have previously been used in the related problems in ECG anomaly detection [31–33]. Meanwhile, LSTM networks perform exceptionally well on sequential data due to their ability to 'remember' what they have previously seen, and thus have previously been trialled in ECG and BP related problems [16,18,34]. By combining the two network structures, we draw on the benefits of both to create a powerful hybrid NN with strong predictive abilities for sequential waveform data. Our proposed hybrid CNN-LSTM outperformed separate CNN and LSTM networks with respect to MAE, SD, and error distribution in our preliminary testing.

As shown in Fig. 1, our network receives a feature vector containing only 5-second raw ECG and PPG waveforms as the input. As ECG and PPG waveforms were both sampled at 125Hz, the feature vector included 625 amplitude data features from both the ECG and PPG waveforms, for a total feature vector size of 1,250.

The proposed network then utilizes three temporal CNN layers, each with 128 hidden units and utilizing ReLU activation. The CNN layers are mathematically described by Eq. (1).

$$y_j^i = relu(\sum_{n=1}^{N} w_{jn}^i * x_m^{(i-1)} + b_j^i) \tag{1}$$

where y_j^i is the jth feature map of the ith layer. Convolution is denoted with the $*$ symbol. Weights $w_{jn}{}^i$ describe the nth weight of the jth feature map from the $(i-1)$th layer, where $n = 1, \ldots, N$. The outputs of the $(i-1)$th layer are denoted as $x_m{}^{(i-1)}$, while bias is denoted as b_j for the jth bias term of the ith layer. Biases are initialised to zero and updated using the Adam optimizer algorithm [35] with a learning rate of 0.01.

Maximum pooling is applied following each convolutional layer, as shown in Fig. 1. Pool1 and Pool2 both use a pool size of 10 and stride of 2, while Pool3 uses a pool size of 4 and stride of 2. Applying maximum pooling after CNN downsamples the outputs, which aids in the prevention of overfitting.

Following the convolutional section of the network, there two bidirectional LSTM network layers with 128 hidden units. Bidirectional LSTMs (BiLSTMs) consider data in both original and reversed order, allowing them to learn from values both in the past and future within the sequence. Results from both forward and reversed sequences are concatenated to provide the overall output, however the mathematical structure for both passes remains the same as standard LSTM. This mathematical process is described Eqs. (2)–(7) below.

$$\tilde{c}_t = \tanh(w_c[\, a_{(t-1)}, x_t\,] + b_c) \tag{2}$$

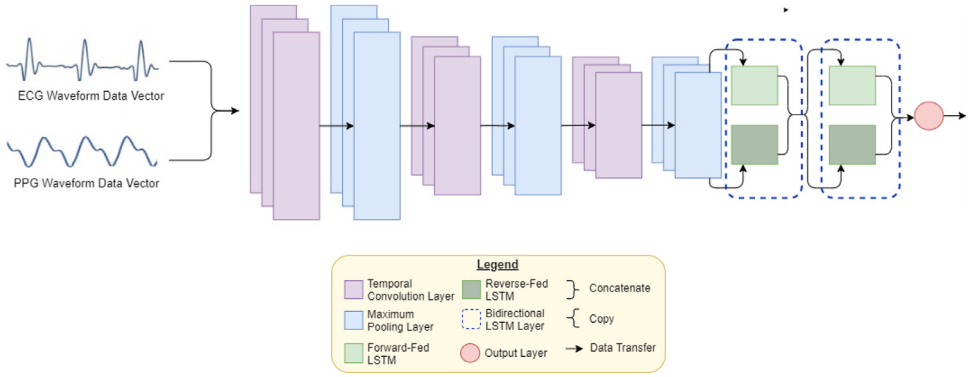

Fig. 1. System model of the proposed DNN for BP estimation.

$$f_t = \sigma\left(w_f[\,a_{(t-1)}, x_t\,] + b_f\right) \tag{3}$$

$$u_t = \sigma\left(w_u[\,a_{(t-1)}, x_t\,] + b_u\right) \tag{4}$$

$$o_t = \sigma\left(w_o[\,a_{(t-1)}, x_t\,] + b_o\right) \tag{5}$$

$$c_t = u_t \bullet \tilde{c}_t + f_t \bullet c_{(t-1)} \tag{6}$$

$$a_t = o_t \bullet \tanh(c_t) \tag{7}$$

where the weights are w_c, w_f, w_u and w_o, while biases are b_c, b_f, b_u and b_o. Biases and weights are learnt using the Adam optimization algorithm [35] with a learning rate of 0.01. The previous layer output is denoted as $a_{(t-1)}$, while x_t is the input to timestep t. Lastly, (6) and (7) are the updated cell state and layer output respectively.

The final layer of the network is a densely-connected node that provides the final output of the network - the SBP or DBP prediction. This network structure was used for training and testing of networks for SBP and DBP estimation separately.

2.5. Training & testing of the DNNs

The DNNs for both SBP and DBP estimation were developed, trained and tested using Keras [36]. The network was trained using 80% of the data for both SBP and DBP networks. A further 10% of the data was used for fine-tuning hyperparameters through validation, with the remaining 10% of data remaining unseen to the networks for testing purposes.

For both the SBP and DBP networks, training and validation was performed over 1,000 epochs with mean absolute error (MAE) used as the loss function. Network performance was checked at the end of each iteration; if the network achieved a lower MAE on the validation set than all previous iterations, then the network weights were saved. If not, then training moved on to the next iteration. The weight combination that resulted in the lowest MAE in the validation set during training was used in the final network, as strong performance on the validation set is indicative of good network fit. Training was conducted on a high-performance laptop with a NVIDIA GeForce GTX 1070 graphics card and took approximately 2.5 h to complete.

Testing was then conducted using the highest-performing networks for SBP and DBP estimation. The results for both networks were recorded and analysed, and are presented with discussion in the following section.

Table 1
Grading criteria defined by the BHS protocol.

	Absolute Difference (mmHg)		
Grade	≤ 5	≤ 10	≤ 15
A	60%	85%	95%
B	50%	75%	90%
C	40%	65%	80%
D	Worse than C		

3. Results & discussion

In evaluating the performance of the proposed CNN-LSTM model for the estimation of SBP and DBP respectively, we consider two widely accepted standards for the approval of blood pressure devices for use in clinical environments - the British Hypertension Society (BHS) protocol and the Association for the Advancement of Medical Instrumentation (AAMI) standard.

Furthermore, we evaluate the level of agreement between the calculations made by the CNN-LSTM networks and the expected SBP and DBP values as determined from ABP waveforms within the MIMIC III database, which were obtained using gold-standard intra-arterial blood pressure measurement. We also compare our work to previous related works, highlighting the improvement that our proposed CNN-LSTM network makes to accurate blood pressure estimation.

Finally, we consider whether the predictions made by the SBP and DBP networks can be combined to produce a prediction for mean arterial pressure (MAP), which represents the average pressure in a person's arteries during a single cardiac cycle [2]. MAP is mathematically defined as follows:

$$MAP = \frac{SBP + (2 \times DBP)}{3}$$

3.1. Comparison to the BHS protocol

The BHS protocol [37] assigns grades of A-D to blood pressure measurement devices, based on the percentages of measurements that achieve absolute differences of less than 5mmHg, 10mmHg, and 15mmHg respectively, when compared to gold-standard measurement techniques such as intra-arterial monitoring. The grading criteria established by the BHS protocol are illustrated in Table 1. Devices that achieve grades of A or B in accordance with the BHS grading criteria are considered suitable for clinical use, while those that achieve lower grades are not recommended for clinical use.

As shown in Table 2, our algorithms satisfy the requirements for an A grade device in the estimation of SBP, DBP, and MAP, and thus would be recommended for use in clinical settings.

Table 2
Assessment of algorithms based on BHS protocol.

	Absolute Difference (mmHg)			
	≤ 5	≤ 10	≤ 15	Grade
SBP	67.66%	89.82%	96.82%	A
DBP	82.79%	96.12%	99.09%	A
MAP	84.21%	97.38%	99.58%	A

Table 3
Assessment of algorithms based on AAMI standard.

	MAE (mmHg)	SD (mmHg)	Grade
SBP	4.41	6.11	Pass
DBP	2.91	4.23	Pass
MAP	2.77	3.88	Pass

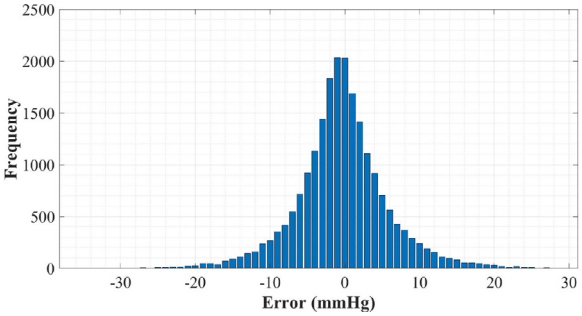

Fig. 2. Error histogram for SBP.

3.2. Comparison to the AAMI protocol

Blood pressure devices are often evaluated with respect to both the AAMI and BHS protocols, as they have different mechanisms for determining device suitability. The AAMI standard [38] states that a device must have a mean difference (also known as mean absolute error) of ≤5 mmHg and a standard deviation (SD) of ≤8 mmHg from gold standard measurements. Devices are assigned a grade of "Pass" if the aforementioned criteria are met, otherwise the device is given a grade of "Fail".

As illustrated in Table 3, our algorithms comfortably achieve "Pass" grades with respect to the AAMI criteria. Our algorithms achieve impressively low MAEs and SD in estimation of all BP parameters, and would be suitable for implementation in healthcare.

3.3. Analysis of error distribution

Accurate measurement of BP is of vital importance in healthcare applications. To further analyse the performance of our proposed CNN-LSTM models, error histograms were generated for SBP, DBP and MAP to inspect the spread of errors.

Figs. 2, 3, 4 present the error distributions for the SBP, DBP, and MAP measurements. Errors are defined as the difference between the true and predicted values for SBP, DBP, and MAP in this context. As the network is a regressor, errors are rounded to the nearest whole number in order to generate discretized error distribution histograms. Each of these histograms clearly shows that 0 mmHg is the most common error, and that most other errors are also extremely low. This in turn indicates a strong level of agreement between the true and predicted values for each BP parameter.

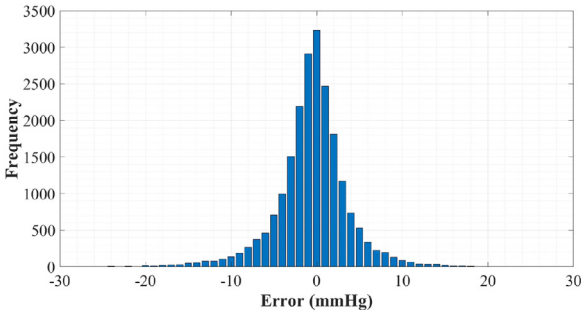

Fig. 3. Error histogram for DBP.

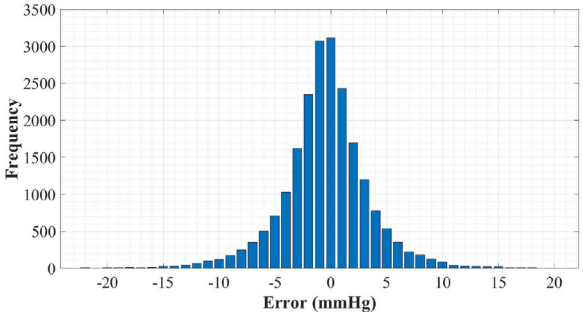

Fig. 4. Error histogram for MAP.

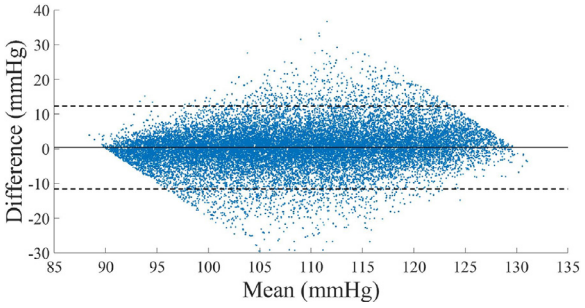

Fig. 5. Bland Altman plot for SBP.

3.4. Level of agreement between intra-arterial monitoring and our CNN-LSTM networks

Bland Altman plots are a key method for assessing the level of agreement between two methods of measurement, particularly in medical applications. These plots illustrate the difference between two measurements compared to the mean of the two measurements, and as such a high density of data near the central 'mean difference' line and between the outer 'limits of agreement' (LOAs) indicates a strong level of agreement between measurements.

In Figs. 5, 6, 7, we compare the SBP, DBP, and MAP predictions made by our CNN-LSTM model with the values obtained using the current gold-standard of BP monitoring, intra-arterial measurement. Each figure clearly shows a high density of points near the mean difference line. The majority of high-range errors are seen around central values for mean as this is the region where the highest level of disagreement between the true and predicted values can occur. The percentage of all points that fell within the LOAs were 93.55%, 93.62% and 93.72% for SBP, DBP, and MAP, respectively. As such, it is clear that there is a high level of agreement between our CNN-LSTM models the respective measurements made via intra-arterial monitoring.

To further evaluate the level of agreement between our proposed scheme and intra-arterial BP measurement, regression plots

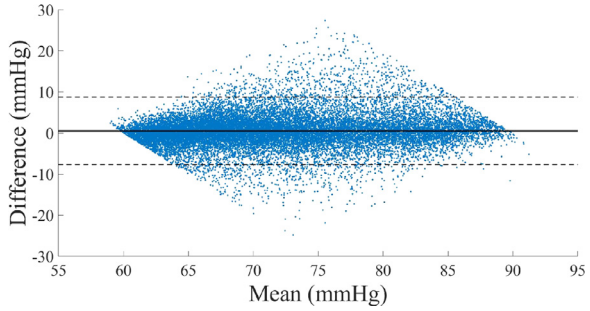

Fig. 6. Bland Altman plot for DBP.

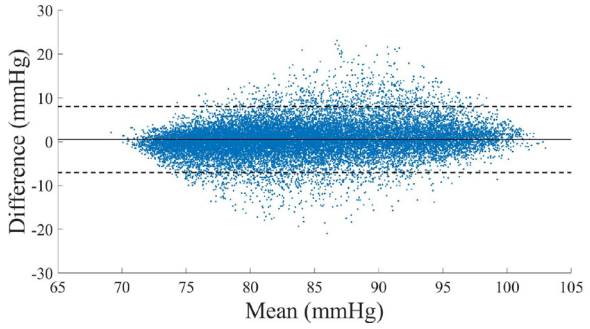

Fig. 7. Bland Altman plot for MAP.

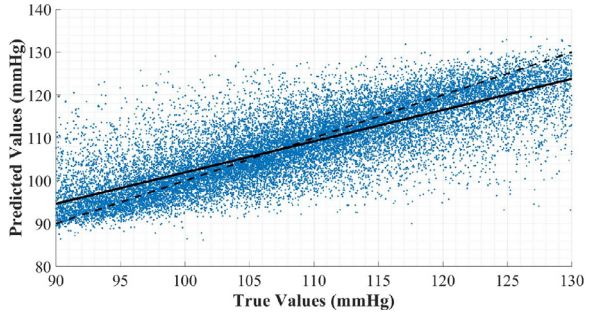

Fig. 8. Regression plot for SBP.

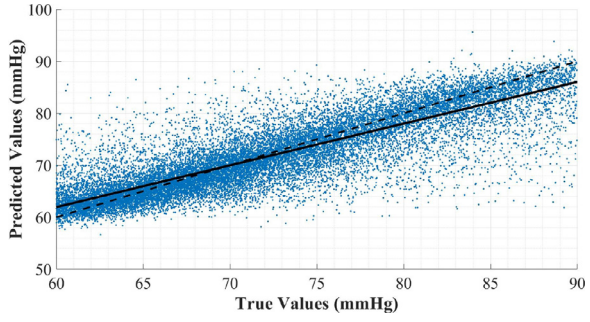

Fig. 9. Regression plot for DBP.

were generated and the coefficients of correlation were calculated to quantify the strength of the relationship between measurements. In all regression plots, the dashed black line shows the theoretical "perfect" correlation, while the solid black line represents the actual correlation. The regression plots for SBP, DBP, and MAP are shown in Figs. 8, 9 and 10 respectively.

Each regression plot illustrates strong positive linear correlation between the "true" values and the actual predictions generated by the CNN-LSTM model. The calculated correlation lines fall close to ideal correlation lines in all cases. To further analyse correlation,

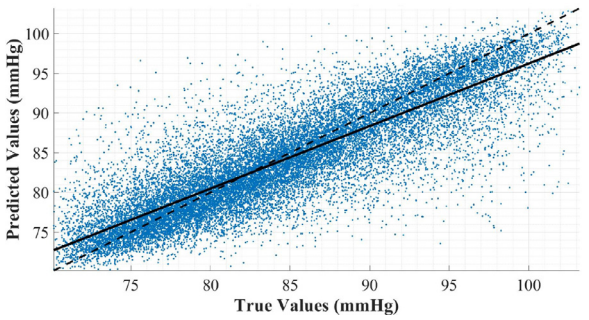

Fig. 10. Regression plot for MAP.

Table 4
Coefficients of correlation.

Blood Pressure Parameter	Coefficient of Correlation
SBP	0.80
DBP	0.85
MAP	0.86

the correlation coefficients were calculated, with the results displayed in Table 4.

These results clearly confirm the strong positive linear relationships between the predictions for SBP, DBP, and MAP made by our scheme when compared with the respective measurements acquired with invasive intra-arterial monitoring.

Overall, it is evident that there is a high level of agreement and strong correlation between our CNN-LSTM networks and the current gold-standard for blood pressure estimation. As our scheme is entirely non-invasive, unlike intra-arterial monitoring, these results are extremely promising for the future of healthcare, especially for at-risk patients such as premature babies and the elderly.

3.5. Comparison to previous works

Several works that are strongly related to our own are presented in [10,12–14,19,20]. Each of these methods utilises PPG and ECG signals to predict blood pressure, though preprocessing and segment lengths vary. Several of these works utilise the MIMIC-II database [12,19,39] or MIMIC III database [20], while others utilise small databases that were independently acquired [13,14]. In all cases, the same database was used for both training and testing of the modes. In this section, we compare our results with the best results achieved by these works. The work in [20] presented results from several databases, and as such we include the results that they obtained using MIMIC III data, as this is the most directly comparable to our own work.

In Table 5, the models developed in previous works are compared to our proposed CNN-LSTM network with respect to the BHS protocol. In terms of SBP, this table shows that our CNN-LSTM network outperforms the previous state-of-the-art works. Our work is the only one to have achieved an 'A' grade for SBP estimation across all data analysed.

Our CNN-LSTM algorithm also performs extremely well for DBP estimation when compared to the previous works. As shown in Table 5, our CNN-LSTM definitively outperforms the DBP prediction networks presented in other works. It also performed favourably to the network presented in [12]. While the latter achieved a higher percentage of results with $\leq$5 mmHg error, our CNN-LSTM had a higher percentage of errors $\leq$10 mmHg and $\leq$15 mmHg. This indicates that the work in [12] has overfit to the data. Our network has not suffered from overfitting, and therefore generates less extremely high errors. While all schemes for DBP estimation achieved grades of 'A' and therefore could be recommended for use by the

Table 5
Comparison of MAP estimation based on the BHS protocol.

		Absolute Difference (mmHg)			
		≤**5**	≤**10**	≤**15**	Grade
Kachuee [10]	SBP	34.1%	56.5%	72.7%	D
	DBP	62.7%	87.1%	95.7%	A
	MAP	54.2%	81.8%	93.1%	B
Mousavi [12]	SBP	71%	77%	84%	C
	DBP	84%	92%	97%	A
	MAP	79%	83%	93%	B
Miao [13]	SBP	51%	81%	94%	B
	DBP	62 %	92%	99%	A
	MAP	60%	90%	98%	A
Song [14]	SBP	N/A	N/A	N/A	B
	DBP	N/A	N/A	N/A	A
	MAP	-	-	-	-
Miao [20]	SBP	50.07%	76.41%	90.39%	B
	DBP	65.66%	89.77%	96.63%	A
	MAP	65.14%	89.58%	96.61%	A
This work	SBP	67.66%	89.82%	96.82%	A
	DBP	82.79%	96.12%	99.09%	A
	MAP	84.21%	97.38%	99.58%	A

Table 6
Comparison of schemes based on the AAMI standard.

		Error Metrics (mmHg)		
		MAE	SD	Grade
Kachuee [10]	SBP	11.80	9.88	Fail
	DBP	5.83	5.71	Fail
	MAP	5.92	5.25	Fail
Mousavi [12]	SBP	3.97	8.90	Fail
	DBP	2.43	4.17	Pass
	MAP	2.61	4.91	Pass
Miao [13]	SBP	6.13	7.76	Fail
	DBP	4.54	5.52	Pass
	MAP	4.81	6.03	Pass
Song [14]	SBP	4.8	6.0	Pass
	DBP	4.8	6.0	Pass
	MAP	N/A	N/A	N/A
Sharifi [19]	SBP	7.83	9.1	Fail
	DBP	4.86	5.21	Pass
	MAP	N/A	N/A	N/A
Miao [20]	SBP	7.10	9.99	Fail
	DBP	4.61	6.29	Pass
	MAP	4.66	6.36	Pass
This work	SBP	4.41	6.11	Pass
	DBP	2.91	4.23	Pass
	MAP	2.77	3.88	Pass

BHS, ours has achieved the lowest number errors greater than 10 mmHg and thus would be considered the most suitable for clinical use.

Our scheme for MAP estimation based is also compared to previous works in Table 5. No results for MAP were presented by [14], but this important diagnostics parameter was examined in [10,12,13,20]. As shown in Table 5, our scheme for MAP estimation outperforms the previous works. With 99.58% of errors falling under 15 mmHg, it is clear that our scheme has very few high-range errors when compared to previous state-of-the-art works. Only our proposed scheme and those of [13,20] achieve the grade of 'A', however our scheme has significantly fewer high-range errors exceeding 5 mmHg and thus would be the most suitable for clinical use.

As shown in Table 6, our work also compares favourably with previous works with respect to the AAMI standard [38]. A common misconception in the literature is the calculation of mean error (ME), rather than mean difference (also known as MAE), when considering the AAMI standard. As such, Table 6 only includes comparisons to papers that provided the correct metric of MAE. In all cases, the grade was determined based on MAE and standard deviation (SD).

Table 6 shows that our CNN-LSTM scheme and that in [14] are the only ones to achieve a grade of "Pass" for SBP estimation, however our scheme has a lower MAE and SD than that of [14]. Our scheme has a marginally higher MAE and SD than the scheme in [12] for DBP, however this is likely due to the overfitting seen in Table 5 for [12]. Based on this, it is clear that our network is more suitable as it performed well on a set-aside testing set, and therefore is highly likely to generalize well to new data.

Finally, the MAE for MAP is slightly higher than the scheme in [12], but the lower SD again indicates that the algorithm has fit better to the data and performs better across all values. It is again likely that the work in [12] is suffers from overfitting to the data. This is evident due to low MAE, but high percentages of high-range errors as shown in Table 5.

It is also worth noting that several recent works, including [15–17], have achieved strong results in terms of parameters such as RMSE and correlation. The work presented by [15] also achieves an A grade in terms of the BHS protocol. However, these works are developed on extremely small datasets comprised of 39 [15], 84 [16], and 37 [17] patients, respectively. Meanwhile, the AAMI criteria require a minimum of 85 participants to be used in testing to ensure validation on a diverse cross-section of the population. While valid pilot studies, these three works would receive immediate grades of "fail" under the AAMI protocol due to insufficient validation of their algorithms at this stage.

Overall, our CNN-LSTM algorithms for SBP, DBP, and MAP perform strongly and have been shown to generalize well to new data. For SBP, DBP and MAP, our proposed model achieved 'A' and "Pass" grades for the BHS and AAMI standard respectively. Ours was the only scheme to meet these standards for all three BP metrics. Additionally, ours was the only scheme to achieve both 'A' and "Pass" grades for SBP measurement. These results suggest that our proposed CNN-LSTM model is the most suitable algorithm for non-invasive BP estimation from ECG and PPG waveforms, and could readily be implemented into healthcare environments.

4. Conclusion

In this work, it was found that hybridized CNN-LSTM networks are capable of estimating key BP parameters from raw ECG and PPG waveforms. Through the use of raw waveforms rather than manual feature selection, we avoid issues associated with time-dependent variables such as PTT, minimise pre-processing, and avoid human bias. Our proposed CNN-LSTM model is capable of accurately estimating SBP, DBP and MAP with high accuracy.

When compared to standards set by the reputable healthcare bodies of BHS and AAMI, our networks performed extremely well. The schemes for estimating SBP, DBP and MAP all met the requirements set by AAMI and achieved grades of 'A' in accordance with the BHS protocol. This success ensures that a device implementing our schemes would be recommended for clinical use by these professional bodies.

Furthermore, when we compared our works to previous state-of-the-art schemes in the literature, our proposed CNN-LSTM was clearly the strongest scheme. Previous schemes performed well in measurement of certain BP parameters, but not in others. Additionally, overfitting was apparent in some schemes. Meanwhile, our scheme is shown to have fit the data extremely well while performing strongly on new data. Additionally, our proposed model is the only scheme that achieves grades of 'A' for the BHS protocol and 'pass' for the AAMI standard across all BP measurements.

These results suggest that our scheme is comparable to sphygmomanometers and other devices used in healthcare environments today, and is certainly ready for clinical trials. In our future work,

we aim to develop hardware with PPG and ECG sensing capabilities that will implement the proposed algorithm, and thereafter undertake clinical trials that include patients with a wider range of blood pressures. This will allow us to validate that the model performs strongly on real-time sensor data and can adapt to handle extremely high or low blood pressure readings.

Overall, the performance of our algorithm indicates that hybrid CNN-LSTM networks are highly suitable for blood pressure prediction. Implementing our algorithm into healthcare monitoring devices would result in non-invasive, continuous, and highly accurate blood pressure measurements for a number of applications, from telehealth to intensive care.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Acknowledgements

This work was supported by the Australian Government Research Training Program Scholarship.

References

[1] World Health Organization, A Global Brief on Hypertension, Tech. rep., World Health Organization, Geneva, Switzerland, 2013.
[2] L. Bonsall, Calculating the mean arterial pressure (MAP), 2011. https://www.nursingcenter.com/ncblog/december-2011/calculating-the-map.
[3] E.A. Wehrwein, M.J. Joyner, Regulation of blood pressure by the arterial baroreflex and autonomic nervous system, in: R.M. Buijs, D.F.B.T.H.o. N. Swaab (Eds.), Handbook of Clinical Neurology, Vol. 117, Elsevier, 2013, pp. 89–102. DOI: 10.1016/B978-0-444-53491-0.00008-0
[4] S. Romagnoli, Z. Ricci, D. Quattrone, et al., Accuracy of invasive arterial pressure monitoring in cardiovascular patients: an observational study, Crit. Care 18 (6) (2014) 644, doi:10.1186/s13054-014-0644-4.
[5] E.M. Frese, A. Fick, S.H. Sadowsky, Blood pressure measurement guidelines for physical therapists, Cardiopulm. Phys. Ther. J. 22 (2) (2011) 5–12, doi:10.1097/01823246-201122020-00002.
[6] Y. Cemal, A. Pusic, B.J. Mehrara, Preventative measures for lymphedema: separating fact from fiction, J. Am. Coll. Surg. 213 (4) (2011) 543–551, doi:10.1016/j.jamcollsurg.2011.07.001.
[7] T.H. Wu, G.K.H. Pang, E.W.Y. Kwong, Predicting systolic blood pressure using machine learning, in: 7th International Conference on Information and Automation for Sustainability, 2014, pp. 1–6. DOI: 10.1109/ICIAFS.2014.7069529
[8] S. Lee, J.H. Chang, Deep Boltzmann regression with mimic features for oscillometric blood pressure estimation, IEEE Sens. J. 17 (18) (2017) 5982–5993, doi:10.1109/JSEN.2017.2734104.
[9] S. Lee, J. Chang, Oscillometric blood pressure estimation based on deep learning, IEEE Trans. Ind. Inf. 13 (2) (2017) 461–472, doi:10.1109/TII.2016.2612640.
[10] M. Kachuee, M.M. Kiani, H. Mohammadzade, et al., Cuffless blood pressure estimation algorithms for continuous health-care monitoring, IEEE Trans. Biomed. Eng. 64 (4) (2017) 859–869, doi:10.1109/TBME.2016.2580904.
[11] Y.Z. Yoon, J.M. Kang, Y. Kwon, et al., Cuff-less blood pressure estimation using pulse waveform analysis and pulse arrival time, IEEE J. Biomed. Health Inf. 22 (4) (2018) 1068–1074, doi:10.1109/JBHI.2017.2714674.
[12] S.S. Mousavi, M. Firouzmand, M. Charmi, et al., Blood pressure estimation from appropriate and inappropriate PPG signals using a whole-based method, Biomed. Signal Process. Control 47 (2019) 196–206, doi:10.1016/j.bspc.2018.08.022.
[13] F. Miao, Z.D. Liu, J.K. Liu, et al., Multi-sensor fusion approach for cuff-less blood pressure measurement, IEEE J. Biomed. Health Inf. 24 (1) (2020) 79–91, doi:10.1109/JBHI.2019.2901724.
[14] K. Song, K. Chung, J. Chang, Cuff-less deep learning-based blood pressure estimation for smart wristwatches, IEEE Trans. Instrum. Meas. (2019), doi:10.1109/tim.2019.2947103. 1–1
[15] M.S. Tanveer, M.K. Hasan, Cuffless blood pressure estimation from electrocardiogram and photoplethysmogram using waveform based ANN-LSTM network, Biomed. Signal Process. Control 51 (2019) 382–392, doi:10.1016/j.bspc.2019.02.028. https://www.sciencedirect.com/science/article/pii/S1746809419300722
[16] P. Su, X.R. Ding, Y.T. Zhang, et al., Long-term blood pressure prediction with deep recurrent neural networks, in: 2018 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2018, Vol. 2018-Janua, 2018, pp. 323–328. DOI: 10.1109/BHI.2018.8333434
[17] A. Paviglianiti, V. Randazzo, E. Pasero, et al., Noninvasive arterial blood pressure estimation using ABPNet and VITAL-ECG, in: 2020 IEEE Int. Instrum. Meas. Technol. Conf., 2020, pp. 1–5, doi:10.1109/I2MTC43012.2020.9129361.
[18] F.P. Lo, C.X. Li, J. Wang, et al., Continuous systolic and diastolic blood pressure estimation utilizing long short-term memory network, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2017, pp. 1853–1856. DOI: 10.1109/EMBC.2017.8037207
[19] I. Sharifi, S. Goudarzi, M.B. Khodabakhshi, A novel dynamical approach in continuous cuffless blood pressure estimation based on ECG and PPG signals, Artif. Intell. Med. 97 (2019) 143–151, doi:10.1016/j.artmed.2018.12.005.
[20] F. Miao, B. Wen, Z. Hu, et al., Continuous blood pressure measurement from one-channel electrocardiogram signal using deep-learning techniques, Artif. Intell. Med. (2020) 101919, doi:10.1016/j.artmed.2020.101919.
[21] A. Goldberger, L. Amaral, L. Glass, et al., The MIMIC-III waveform database, 2016. https://archive.physionet.org/physiobank/database/mimic3wdb/.
[22] Samsung, Samsung health monitor, 2021. https://www.samsung.com/us/apps/samsung-health-monitor/.
[23] Apple, Healthcare, 2021, https://www.apple.com/healthcare/apple-watch/.
[24] Y. Shahriari, R. Fidler, M.M. Pelter, et al., Electrocardiogram signal quality assessment based on structural image similarity metric, IEEE Trans. Biomed. Eng. 65 (4) (2018) 748–753, doi:10.1109/TBME.2017.2717876.
[25] A.E. Johnson, T.J. Pollard, L. Shen, et al., MIMIC-III, a freely accessible critical care database, Sci. Data 3 (2016) 160035, doi:10.1038/sdata.2016.35.
[26] J. Li, H.W. Lewis, Fuzzy clustering algorithms - review of the applications, in: Proceedings - 2016 IEEE International Conference on Smart Cloud, SmartCloud 2016, 2016, pp. 282–288. DOI:10.1109/SmartCloud.2016.14
[27] Y. Zhang, S. Wei, L. Zhang, et al., A signal quality assessment method for mobile ECG using multiple features and fuzzy support vector machine, in: 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, ICNC-FSKD 2016, 2016, pp. 966–971. DOI: 10.1109/FSKD.2016.7603309
[28] C. Orphanidou, T. Bonnici, P. Charlton, et al., Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring, IEEE J. Biomed. Health Inf. 19 (3) (2015) 832–838, doi:10.1109/JBHI.2014.2338351.
[29] J. Seladi-Schulman, C. Stephens, Pulse pressure calculation explained, 2017. https://www.healthline.com/health/pulse-pressure.
[30] S.G. Sheps, Pulse pressure: an indicator of heart health? - Mayo clinic, 2016. Https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/expert-answers/pulse-pressure/faq-20058189%0Ahttp://www.mayoclinic.org/diseases-conditions/high-blood-pressure/expert-answers/pulse-pressure/faq-20058189.
[31] X. Zhai, C. Tin, Automated ECG classification using dual heartbeat coupling based on convolutional neural network, IEEE Access 6 (2018) 27465–27472.
[32] X. Fan, Q. Yao, Y. Cai, et al., Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings, IEEE J. Biomed. Health Inf. 22 (6) (2018) 1744–1753, doi:10.1109/JBHI.2018.2858789.
[33] S. Kiranyaz, T. Ince, M. Gabbouj, Real-time patient-specific ECG classification by 1-D convolutional neural networks, IEEE Trans. Biomed. Eng. 63 (3) (2016) 664–675, doi:10.1109/TBME.2015.2468589.
[34] S. Chauhan, L. Vig, Anomaly detection in ECG time signals via deep long short-term memory networks, in: Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015, 2015, pp. 1–7. DOI: 10.1109/DSAA.2015.7344872
[35] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization. ArXiv preprint arXiv:1412.6980.
[36] F. Chollet, et al., Keras, 2015. https://keras.io.
[37] E. O'Brien, J. Petrie, W. Littler, et al., Short report: an outline of the revised british hypertension society protocol for the evaluation of blood pressure measuring devices, J. Hypertens. 11 (6) (1993) 677–679, doi:10.1097/00004872-199306000-00013.
[38] Association for the Advancement of Medical Instrumentation, The national standard for electronic or automated sphygmomanometers, AAMI, Arlington, VA, 1987.
[39] M. Kachuee, M.M. Kiani, H. Mohammadzade, et al., Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time, in: Proceedings - IEEE International Symposium on Circuits and Systems, Vol. 2015-July, 2015, pp. 1006–1009. DOI: 10.1109/ISCAS.2015.7168806