Variation of the Korotkoff Stethoscope Sounds During Blood Pressure Measurement: Analysis Using a Convolutional Neural Network

Fan Pan [ID], Peiyu He, Chengyu Liu, Taiyong Li, Alan Murray, and Dingchang Zheng [ID]

Abstract—**Korotkoff sounds are known to change their characteristics during blood pressure (BP) measurement, resulting in some uncertainties for systolic and diastolic pressure (SBP and DBP) determinations. The aim of this study was to assess the variation of Korotkoff sounds during BP measurement by examining all stethoscope sounds associated with each heartbeat from above systole to below diastole during linear cuff deflation. Three repeat BP measurements were taken from 140 healthy subjects (age 21 to 73 years; 62 female and 78 male) by a trained observer, giving 420 measurements. During the BP measurements, the cuff pressure and stethoscope signals were simultaneously recorded digitally to a computer for subsequent analysis. Heartbeats were identified from the oscillometric cuff pressure pulses. The presence of each beat was used to create a time window (1 s, 2000 samples) centered on the oscillometric pulse peak for extracting beat-by-beat stethoscope sounds. A time-frequency two-dimensional matrix was obtained for the stethoscope sounds associated with each beat, and all beats between the manually determined SBPs and DBPs were labeled as "Korotkoff." A convolutional neural network was then used to analyze consistency in sound patterns that were associated with Korotkoff sounds. A 10-fold cross-validation strategy was applied to the stethoscope sounds from all 140 subjects, with the data from ten groups of 14 subjects being analyzed separately, allowing consistency to be evaluated between groups. Next, within-subject variation of the Korotkoff sounds analyzed from the three repeats was quantified, separately for each stethoscope sound beat. There was consistency between folds with no significant differences between groups of 14 subjects ($P = 0.09$ to $P = 0.62$). Our results showed that 80.7% beats at SBP and 69.5% at DBP were analyzed as Korotkoff sounds, with significant differences between adjacent beats at systole (13.1%, $P = 0.001$) and diastole (17.4%, $P < 0.001$). Results reached stability for SBP (97.8%, at sixth beat below SBP) and DBP (98.1%, at sixth beat above DBP) with no significant differences between adjacent beats (SBP $P = 0.74$; DBP $P = 0.88$). There were no significant differences at high-cuff pressures, but at low pressures close to diastole there was a small difference (3.3%, $P = 0.02$). In addition, greater within subject variability was observed at SBP (21.4%) and DBP (28.9%), with a significant difference between both ($P < 0.02$). In conclusion, this study has demonstrated that Korotkoff sounds can be consistently identified during the period below SBP and above DBP, but that at systole and diastole there can be substantial variations that are associated with high variation in the three repeat measurements in each subject.**

Index Terms—**Blood pressure, convolutional neural network, Korotkoff sound.**

I. INTRODUCTION

NON-INVASIVE blood pressure (BP) measurements are commonly used by professional healthcare providers and the general public at home. There are two main non-invasive BP measurement techniques: manual auscultatory and automatic oscillometric techniques. Manual auscultatory BP measurement has been in use for over one hundred years, and is currently the most accurate non-invasive BP measurement and is considered the gold standard clinical BP measurement [1]. Since manual auscultatory BP measurement requires professional training before the operator is competent to perform the measurement, automatic BP devices have been widely used because they are easy to operate and require less training [2].

Manual auscultatory is achieved by wrapping a non-distensible cuff around the arm, and placing a stethoscope over the brachial artery. The cuff is inflated by the operator until the cuff pressure is above the brachial artery pressure to block blood flow. Then the cuff pressure is released linearly and slowly (2–3 mmHg/s), and during cuff deflation a series of sounds associated with each heartbeat between systole and diastole can be heard, and they are named Korotkoff sounds. The first Korotkoff sound heard determines SBP. With further cuff pressure deflation, the

artery opens and closes, causing more turbulent blood flow and further Korotkoff sounds to be heard. The last Korotkoff sound heard before complete disappearance of the sounds determines DBP, and is associated with smooth blood flow [3], [4]. The most important point for manual auscultatory measurement is that Korotkoff sounds need to be identified correctly. However, there are challenges for accurate identification, due to the changing of characteristics of Korotkoff sounds during BP measurement, between different measurements or between different subjects. These can cause uncertainties for SBP and DBP determination [5]–[7].

Recently, deep learning techniques have been applied to various research fields with impressive outcomes, such as for computer vision, automatic speech recognition and natural language processing [8]–[10]. This raises the expectation that deep learning can be extended to medical fields. Unlike traditional machine learning, deep learning techniques have multiple layers of non-linear processing and do not rely on feature selection to obtain reliable results. They can automatically detect and generate multi-dimension features from raw data sources. Among them, convolutional neural network (CNN) has been regarded as an attractive deep learning architecture. CNN has been widely used as classifiers for image processing with good performance of binary or multi-class classification. Some researchers have demonstrated its effectiveness and verified that it can beat the traditional methods for the applications of brain image segmentation, ECG classification, and lung tissue characterization [11]–[13]. So, CNN has therefore the potential for identifying the characteristics change of Korotkoff sounds.

In this paper, a specific CNN structure was designed to identify the Korotkoff sound beats from stethoscope sound. Then the consistency in sound patterns that were associated with Korotkoff sounds, as well as the within-subject variation in the repeat measurements were analyzed by examining all stethoscope sounds associated with each heartbeat from above systole to below diastole during linear cuff deflation.

II. METHODS

A. Subjects

One hundred and forty healthy subjects (62 female and 78 male) were enrolled in this study, aged from 21 to 73 years (42 $\pm$ 12 years). All subjects gave their written informed consent to participate. None had any known cardiovascular disease. The study received ethical permission from the Newcastle & North Tyneside Research Ethics Committee. The investigation conformed with the principles in the Declaration of Helsinki. The detailed subject demographic information, including age, sex, height, weight and arm circumference, is summarized in Table I. All the analysis involved was performed on anonymised data.

B. BP Measurement

Three repeat BP measurements were taken from each subject, with a four-minute interval between the repeat measurements, allowing recovery of cardiovascular hemodynamics. All BP measurements were performed in a quiet and

TABLE I
GENERAL DATA INFORMATION FOR THE SUBJECTS STUDIED

		Min	Max	Mean	SD
No. subjects	140				
No. male	78				
No. female	62				
Age (years)		21	73	42	12
Height (cm)		152	196	172	10
Weight (kg)		39	108	74	13
Arm circumference (cm)		22	39	29	3

temperature-controlled clinical measurement room. Prior to the measurement, each subject was asked to rest on a chair for 5 minutes and to breathe gently during the whole measurement. The whole measurement procedure followed the guidelines recommended by the British Hypertension Society and American Heart Association [1], [14].

During each BP measurement, the cuff was firstly inflated and then deflated linearly at the recommended rate of 2–3 mmHg/s. Manual auscultatory SBP and DBP were determined from the appearance and disappearance of the Korotkoff sounds by a trained operator, using a clinically validated manual electronic sphygmomanometer (Accoson Greenlight 300; AC Cossor & Son (Surgical) Ltd, Harlow, UK) to display the pressure in the cuff. During cuff deflation, the cuff pressure was recorded by a pressure sensor connected to the cuff via a tube, and the stethoscope sounds were simultaneously recorded by a bespoke system that included a stethoscope end and a microphone. Then the signals were analogue-to-digital converted with a sampling rate of 2000 Hz and resolution of 16 bits. The final digital signals were stored in a computer for off-line analysis. A total of 420 stethoscope recordings (from 3 repeat measurements × 140 subjects) were obtained in this study.

C. Stethoscope Signal Processing

The linearly deflating baseline cuff pressure was firstly removed to obtain the oscillometric pulses (OscP). All the peaks of OscP associated with heartbeats were detected and used as reference points to segment the stethoscope signal into second-by-second stethoscope sound frames (1s window with 2,000 sample points per frame, centered with the peak at each OscP). Fig. 1 illustrates an example for the stethoscope signal segmentation. The left arrow in the figure points to a segmented non-Korotkoff stethoscope sound frame and the other arrow points to a segmented Korotkoff stethoscope sound frame.

After the stethoscope signal was segmented, each frame was converted into matrix 'images' by Short Time Fourier Transformation (STFT) with 60 ms Hamming window (sampling rate = 2000 Hz) and 87% overlap. As shown in Fig. 2, the x-axis represents time, y-axis is frequency, and the value of every pixel indicates the power at a particular time and frequency. The converted matrix image was 116 (time) × 60 (frequency) pixel points, so its resolution in time and frequency domain was about 9 ms and 17 Hz respectively.
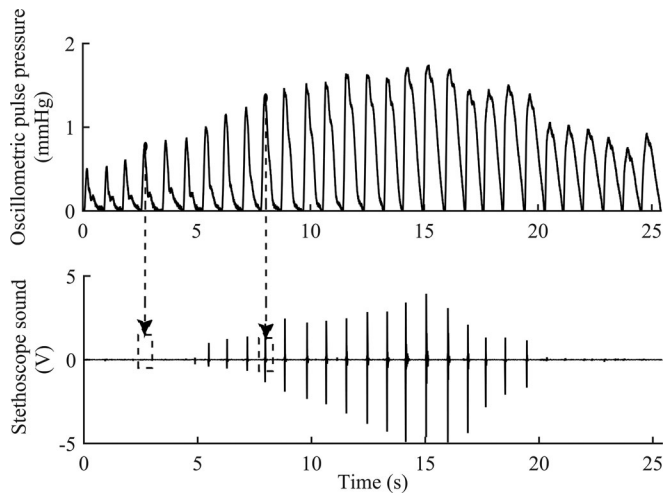
Fig. 1.   Illustration of stethoscope signal segmentation. Peaks of os-cillometric cuff pressure pulses (OscP) were detected for stethoscope signal segmentation. Stethoscope signal was segmented into second-by-second frames (2,000 sample points per frame, centered with the peak of OscP).
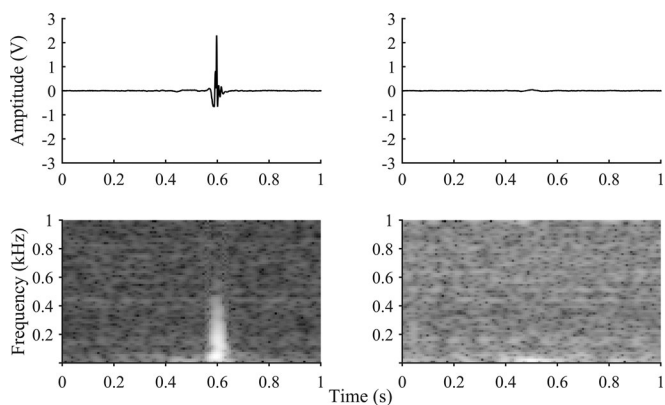

Fig. 2.   Korotkoff (left) and non-Korotkoff (right) sound frames, and the converted time-frequency transformation images.

D. Manual Labeling of Korotkoff and Non-Korotkoff Sound Beats

Each segmented stethoscope sound frame was manually la-beled as 'Korotkoff' or 'non-Korotkoff' sound beats based on the reference manual auscultatory BP results, i.e., the frames corresponding with the stethoscope signal between the manu-ally determined SBP and DBP were labeled as Korotkoff sound beats, and the frames above manual SBP or below DBP were labeled as non-Korotkoff sound beats.

E. Korotkoff Sound Beat Identification Using Convolutional Neural Network

A deep learning based approach, convolutional neural net-work (CNN), was used to analyze consistency in sound patterns that were associated with Korotkoff sounds. CNN is a type of feed-forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of

the visual cortex. Individual cortical neurons respond to stimuli in a restricted region of space known as the receptive field. The response of an individual neuron to stimuli within its recep-tive field can be approximated mathematically by a convolution operation.

Fig. 3 shows the CNN structure designed in this study, which was convolved by three layers. Each layer consisted of a con-volutional and a pooling layer in succession. There were 32 kernels in the first and second convolutional layers and 64 ker-nels in the third convolutional layer. Each kernel was a neuron inside a layer. The convolutional layer $k \times k$, means that the input image (size of $N \times M$) becomes $(N-k + 1) \times (M-k + 1)$ after the convolutional operation. Since the kernel dimension usually has a small size, CNN exploits spatially local correla-tion by enforcing a local connectivity pattern between neurons of adjacent layers. The architecture thus ensures that the learnt "filters" produce the strongest response to a spatially local input pattern. Stacking many such layers leads to non-linear "filters" that become increasingly "global" (i.e., responsive to a larger region of pixel space). This allows the CNN network to first create good representations of small parts of the input, then as-semble representations of larger areas from them. The pooling layer had a size of 2×2, which halved the size of the image after the convolutional operation.

In this study, the input of the network was a 116×60 (N $\times$ M) image patch, the convolutional layer was 5×5 and the pooling layer was 2×2. After the first convolutional layer, the image became 112×56 $((N-k + 1) \times (M-k + 1))$, and then became 56×28 after the first pooling layer. After the second convolutional layer, the image became 52×24, and 26×12 after the second pooling layer. After the third convolutional layer, the image became 22×8, and 11×4 after the third pooling layer. At last, there was a fully connected layer and a logistic regression classifier to provide the identification of Korotkoff or non-Korotkoff sound beats. A ReLU function was used as an activation function in this study to speed up the training process, which was shown as: $f(x) = \max(x, 0)$ (where the input value x represents the output of the upper layer). The CNN process was performed using a computer with Windows Operating System with CPU Intel Xeon E5640 @ 2.66 GHz, GPU NVIDIA GTX 1060.

F. Variation of Stethoscope Sounds Analyzed by Convolutional Neural Network

A 10-fold cross-validation strategy was applied to the stetho-scope sounds from all 140 subjects, with the data from ten groups of 14 subjects being analysed separately, allowing con-sistency to be evaluated between groups. For all the 42 mea-surements within each group, the number (and percentage) of beats identified as Korotkoff sounds was obtained, sep-arately for all beats from above SBP to below DBP dur-ing BP measurement. The variation of Korotkoff sounds dur-ing cuff deflation was investigated by comparing the per-centages of beats analysed as Korotkoff sounds at SBP and DBP with the results from the other beats above and be-low SBP and DBP. The consistency between folds was also
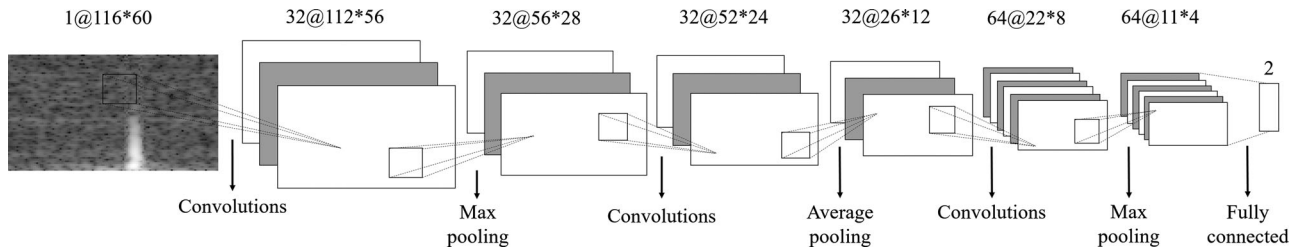
Fig. 3.   Illustration of CNN structure with three layers of convolutional and pooling layers. A fully connected layer and a logistic regression classifier were used to predict if an input image belonged to a Korotkoff or non-Korotkoff sound beat.
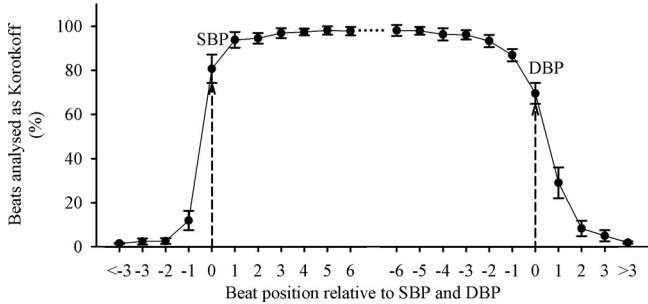

Fig. 4.   Mean percentage of beats analysed as Korotkoff for beats above SBP to beats below DBP. The 95% confidence intervals are given for variation between the ten folds (groups of 14 subjects).

Fig. 5.   Within subject SD (%) for beats above SBP to beats below DBP. The 95% confidence intervals are given for variation between all 140 subjects.

analyzed, with the 95% confidence interval (2xSEM) between folds calculated.

G. Within Subject Variability of Stethoscope Sounds

To evaluate the within subject variability of Korotkoff sound identification, the SD of three repeat measurements from each subject was calculated for all beats. The mean variation (SD) and its confidence interval were then calculated across all 140 subjects, separately for all beats. Minitab 17 software package (Minitab Inc, State College, PA, USA) was employed to perform the analysis. A value of $P < 0.05$ was considered as statistically significant.

III. RESULTS

A. Variation of Analyzed Korotkoff Beats

The variations in identification of Korotkoff sounds during cuff deflation are showed in Fig. 4, for all beats from above SBP to below DBP. In total, there were 7378 Korotkoff sounds and 12852 non-Korotkoff sounds detected. There was an average of 18 Korotkoff beats from each measurement.

There was consistency between folds with no significant differences between groups of 14 subjects ($P = 0.09$ to $P = 0.62$). Our results showed that 80.7% beats at SBP and 69.5% at DBP were analysed as Korotkoff sounds, with significant differences between adjacent beats at systole (13.1%, $P = 0.001$) and diastole (17.4%, $P < 0.001$). Results reached stability for SBP (97.8%, at 6th beats below SBP) and DBP (98.1%, at 6th beat above DBP) with no significant differences between adjacent beats (SBP $P = 0.74$; DBP $P = 0.88$). There were no significant
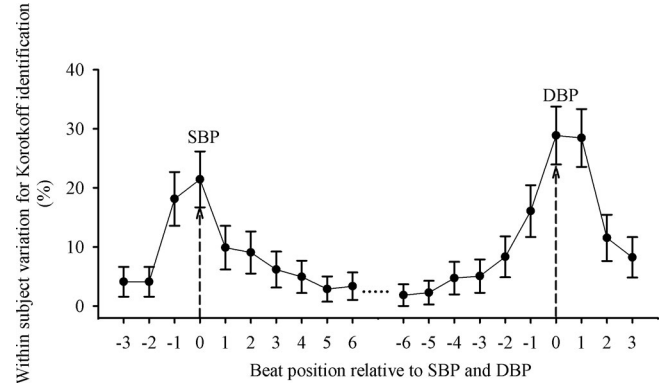
differences at high cuff pressures, but at low pressures close to diastole there was a small difference (3.3%, $P = 0.02$).

B. Within Subject Variability of Stethoscope Sounds

Fig. 5 shows the within subject variation for all beats. Greater within subject variability was observed at manually identified SBP (21.4%) and DBP (28.9%), with the greatest at DBP, and with a significant difference between both ($P < 0.02$). There was smaller variability and more consistent identification results at other beat positions. These results confirm the uncertainties for manual systolic and diastolic pressure (SBP and DBP) determinations.

IV. CONCLUSION

This study quantitatively assessed the variation of Korotkoff sounds from above systole to below diastole during BP measurement, as well as the within-subject variability of Korotkoff sounds between repeats. A deep learning based CNN method has been developed and applied at beat-by-beat level to classify whether stethoscope sounds could be identified as Korotkoff sounds. To the best of our knowledge, this is the first study to apply CNN method to assess the variation of Korotkoff sounds during BP measurement.

Our results provided scientific evidence that, in real clinical practice, the Korotkoff sounds at systole and diastole are much more difficult to be clearly determined than the other beats below SBP and above DBP. It has been widely accepted
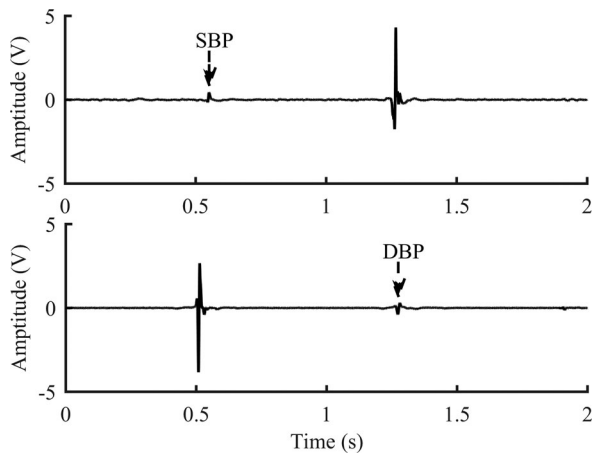
Fig. 6. Waveform example with low quality of Korotkoff sound beats at SBP and DBP. Their amplitudes were much lower than the normal Korotkoff sound beats nearby.

that there are uncertainties for SBP and DBP determinations using manual auscultatory method, and there are differences in BP measurements between operators [15], [16]. These uncertainties are partially caused by the unclear Korotkoff sounds heard by the operator at systole and diastole. During manual auscultatory BP measurement, the first Korotkoff sound heard at SBP correspond to the opening of the artery from collapse with a weak sound. The sound characteristic may not be significantly distinct when compared with the non-Korotkoff sound beats, resulting in relatively low identification accuracy and relatively large within-subject variability. Below SBP, the amplitude of Korotkoff sounds become larger, and their characteristics become reliable, leading to increased identification accuracy. Similarly, when the pressure in the cuff deflates to DBP, the Korotkoff sounds enter phase IV or phase V, and their characteristics become indistinct again, especially at DBP [17]. Therefore, our quantitative results (around 80–85% beats analysed as Korotkoff sounds in comparison with nearly 100% at 5 beats below SBP) agreed with the physiological changes of Korotkoff sounds characteristics during BP measurement. Fig. 6 demonstrates a waveform example for low quality SBP beat and DBP beat respectively. It can be observed from the figure that, compared with the normal Korotkoff sound beats, the amplitude of low quality SBP and DBP beats were much lower.

It has been reported that there was variation in manual BP measurement between repeats, with a SD of 3.7 mmHg in SBP and 3.2 mmHg in DBP [18]. It is also suggested that repeated BP measurements should be made to avoid erroneous diagnosis because of the BP variability [1]. Our study quantified the within-subject variation of Korotkoff sound characteristics at the beat-by-beat level during BP measurement. Greater within-subject variability was observed at systole (21.4%) and diastole (28.9%), which decreased to nearly 0% at six beats below SBP, demonstrating that the characteristics at SBP and DBP were not consistent between repeat measurements. This can also be caused by the unclear Korotkoff sounds identified at SBP and DBP, leading to indistinct characteristics at these two beats.

Next, when the CNN method was developed in this study to identify Korotkoff sounds, the time-frequency transformation was applied to convert the 1-D stethoscope signals into 2-D time-frequency images. This was to achieve better accuracy in classifying Korotkoff and non-Korotkoff sound beats, because CNN can extract more characteristics from the 2-D time-frequency images than from the 1-D time serial signal. Additionally, the small variation of Korotkoff sounds identification between the 10 folds demonstrated that the CNN-based method has a reliable binary classification ability to identify the Korotkoff sound beats. With these advantages, the CNN-based Korotkoff sound identification could be developed further in a future study to realize a CNN-based BP measurement technique.

One limitation of our study is that, during the segmentation of stethoscope signal, a window of 1s was used to convert the 1-D stethoscope signal into 2-D time-frequency image. With some extreme situations, such as with the patients with very fast heart rate (>150 beat/min), one frame could include more than two Korotkoff beats, affecting its identification performance. However, this did not happen in this study since there was no subject with a fast heart rate. In the future, a more advanced signal processing approach needs to be developed to avoid its potential effect with fast heart rates. In addition, the effectiveness of the CNN method on some specific cardiovascular patients, such as patients with ectopic beats and atrial fibrillation, needs to be further investigated to confirm its effectiveness.

In conclusion, this study has demonstrated that Korotkoff sounds can be consistently identified during the period below SBP and above DBP, but that at systole and diastole there can be substantial variations that are associated with high variation in the three repeat measurements in each subject.

REFERENCES

[1] G. Beevers, G. Y. Lip, and E. O'Brien, "Blood pressure measurement: Part II–conventional sphygmomanometry: Technique of auscultatory blood pressure measurement," *Brit. Med. J.*, vol. 322, no. 7293, pp. 1043–1047, 2001.

[2] M. Nitzan, "Automatic noninvasive measurement of arterial blood pressure," *IEEE Instrum. Meas. Mag.*, vol. 14, no. 1, pp. 32–37, Feb. 2011.

[3] E. O'Brien *et al.*, "European society of hypertension recommendations for conventional, ambulatory and home blood pressure measurement," *J. Hypertension*, vol. 21, no. 5, pp. 821–848, 2003.

[4] E. T. O'Brien and K. O'Malley, "ABC of blood pressure measurement: Technique," *Brit. Med. J.*, vol. 2, no. 6196, pp. 982–984, 1979.

[5] L. Geddes and A. Moore, "The efficient detection of Korotkoff sounds," *Med. Biol. Eng. Comput.*, vol. 6, no. 6, pp. 603–609, 1968.

[6] C. F. Babbs, "The origin of Korotkoff sounds and the accuracy of auscultatory blood pressure measurements," *J. Amer. Soc. Hypertension*, vol. 9, no. 12, pp. 935–950, 2015.

[7] M. A. Brown, M. L. Buddle, T. Farrell, G. Davis, and M. Jones, "Randomised trial of management of hypertensive pregnancies by Korotkoff phase IV or phase V," *The Lancet*, vol. 352, no. 9130, pp. 777–781, 1998.

[8] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proc. Int. Conf. IEEE Comput. Vis. Pattern Recognit.*, 2012, pp. 3642–3649.

[9] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," in *Proc. Int. Conf. IEEE Acoust., Speech Signal Process.*, 2013, pp. 8599–8603.

[10] G. Mesnil *et al.*, "Using recurrent neural networks for slot filling in spoken language understanding," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 3, pp. 530–539, Mar. 2015.

[11] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. De Vries, M. J. Benders, and I. Išgum, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1252–1261, May 2016.

[12] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.

[13] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1207–1216, May 2016.

[14] T. Pickering *et al.*, "Subcommittee of professional and public education of the American Heart Association Council on High Blood Pressure Research. 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*, vol. 45, no. 1, pp. 142–161, 2005.

[15] P. E. Nielsen, B. Oxenbøll, K. Astvad, and F. Gyntelberg, "Auscultatory measurement of blood pressure performed by the doctor on duty," *Acta Medica Scandinavica*, vol. 198, no. 1/6, pp. 35–37, 1975.

[16] P. D. Neufeld and D. L. Johnson, "Observer error in blood pressure measurement," *CMAJ: Can. Med. Assoc. J.*, vol. 135, no. 6, pp. 633–637, 1986.

[17] J. Allen, T. Gehrke, J. J. O'Sullivan, S. T. King, and A. Murray, "Characterization of the Korotkoff sounds using joint time–frequency analysis," *Physiol. Meas.*, vol. 25, no. 1, pp. 107–117, 2003.

[18] D. Zheng, R. Giovannini, and A. Murray, "Effect of respiration, talking and small body movements on blood pressure measurement," *J. Human Hypertension*, vol. 26, no. 7, pp. 458–462, 2012.