Convolutional Neural Network for ECG-based Virtual Pathology Stethoscope Tracking in Patient Heart Auscultation

Abstract—**Cardiac auscultation (CA) is the auditory detection of heart sounds to diagnose abnormalities, a crucial skill that is both efficient and cost-effective in medical practice. However, due to increased prevalence and usage of expensive cardiac technologies, many new physicians and trainees have difficulty performing essential cardiac examinations on their patients, particularly diagnosing abnormalities through auscultation using a stethoscope. A virtual pathology stethoscope is a simulation-based solution that train students to perform cardiac examinations by listening to abnormal heart sounds in otherwise healthy standardized patients (SPs). This study reports the accuracy of an electrocardiogram (ECG)-based stethoscope tracking method for placing virtual symptoms in correct auscultation regions. A modified stethoscope head with two electrodes was used to pick up ECG signals at the four primary auscultation sites. A one-dimensional convolutional neural network (CNN) is then modeled to classify the location of the stethoscope by taking advantage of subtle differences in the ECG signals. A 91% accuracy was obtained, showing promising performance gain over our previous methods. This finding would significantly extend the simulation capabilities of SPs by allowing trainees to perform realistic CA and hear CA in clinical environment.**

Index Terms—**Electrocardiogram, convolutional neural network, clinical simulation, auscultation.**

I. Introduction

Cardiovascular Disease (CVD) is the leading cause of death in the United States, accounting for 25% of the total death each year [1], [2]. According to a projection done by the American Heart Association [2], by the year 2030, 40% of the U.S. population is expected to suffer from CVD. Furthermore, the number of required primary care and specialty physicians is expected to decline (an estimated shortage of 16,000 cardiologists by 2050). Thus, there is an urgent need prepare more health-care workers for the expected demand.

Cardiac auscultation (CA) is the auditory detection of heart sounds to diagnose abnormalities, a crucial skill that is both efficient and cost-effective in medical practice [3]. However, due to increased prevalence and usage of expensive cardiac technologies, such as echocardiography, many new physicians and trainees have difficulty performing essential cardiovascular examinations on their patients, particularly diagnosing abnormalities through auscultation using a stethoscope [4]–[6]. Because the stethoscope is an inexpensive technology that relies on the physician to accurately locate auscultation sites and diagnose common structural cardiac conditions, trainees should focus on perfecting this skill rather than turning to cost-inefficient technologies that serve a similar function.

The areas of auscultation are generally correlated with the cardiac valves. Proper auscultation requires a consistent approach to site detection. Typically a sequential method is used, assessing four different sites in the following order: (1) upper right sternal border (aortic site), (2) upper left sternal border (pulmonic site), (3) lower left sternal border (tricuspid site), and (4) apex (mitral site) (see Fig. 1) [7], [8].

Many institutions have incorporated standardized patients (SPs) into cardiac auscultation training; because these patients, though actors, can portray patients and communicate their symptoms [9]. SPs are typically healthy individuals, limiting the kinds of abnormalities that students can hear [7]. We are currently developing and testing an electrocardiogram-based tracking stethoscope to simulate a realistic cardiac examination [7], [10]. A stethoscope with ECG-based tracking takes advantage of subtle ECG signal differences among the four auscultation sites to identify which site the stethoscope is placed upon, introducing the potential for playback devices to produce an abnormal heart sound when the site is correctly identified.

Our previous system, using a Random Forest classifier [10], correctly identified the auscultation sites with an accuracy of 86% irrespective of the SP posture (seated or supine), obesity (Body Mass Index), or sex. The accuracy also remained stable for different stethoscope orientation tested by placing the chest-piece at different angles over SP's torso [7], [10]. While our results showed that the method could be extended to a wider subject pool, the accuracy required for an online real-time tracking was found to be sub-optimal.

This paper investigates the performance of our newly-developed prediction algorithm to provide higher accuracy for a real-time application. The proposed system employs a convolutional neural network (CNN) model to detect the four CA sites. We trained the model on the raw signal as well as on characteristic features (e.g. QRS wave) extracted from the single-lead ECG recordings. Fig. 1 presents an overview of the proposed system.

This paper is organized as follows; Section II will summarize related works on virtual pathology stethoscope, while Section III will present the methods for our proposed CNN system. Section IV will highlight experimental results and subsequent model evaluation. Finally, Section V will discuss conclusive remarks and propose potential future work.

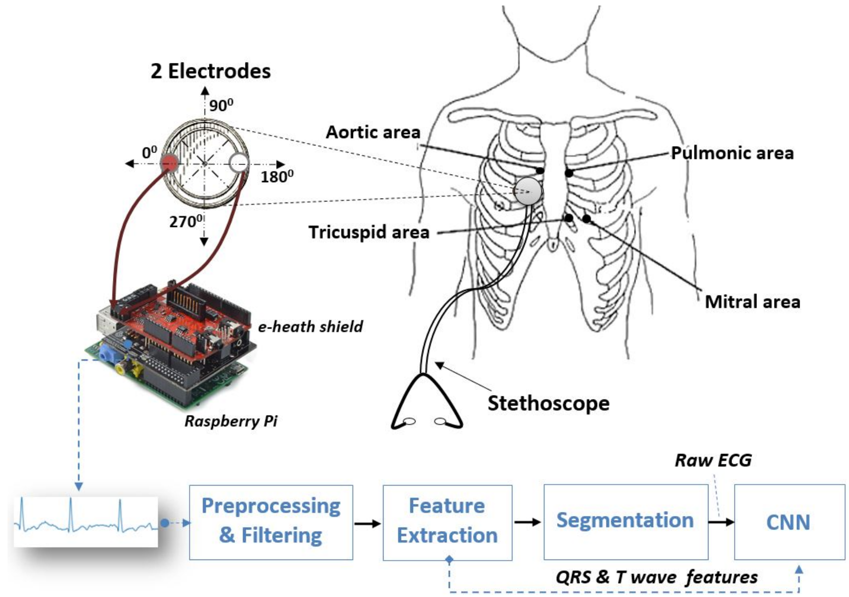

Fig. 1. Stethoscope apparatus and proposed system.(top-right) The four primary cardiac auscultation sites that the algorithm can detect. (top-left) Two direct-contact electrodes fixed on a standard stethoscope head to record ECG signals. (bottom) The setup was attached to an e-health sensor connected to a Raspberry Pi computer to run the ECG signal acquisition program.

II. Related Works

Virtual pathology stethoscope (VPS) technologies that can play abnormal heart sounds during auscultation are being developed to simulate a variety of cardiac conditions. One commercially available instructor or SP triggered VPS is the Ventriloscope [11]. The system uses infrared technology to receive MP3 files from a remote transmitting unit. However, Ventriloscope requires additional SP training to deliver a fully integrated simulation. Also, the transmitter requires a direct line of sight for signaling.

In [12], F. D. McKenzie, H. M. Garcia, and R. J. Castelino designed a VPS that can automatically play abnormal sounds by tracking the location of the stethoscope over the SPâs torso. The system used magnetic tracking technology to find the head of the stethoscope. However, the setup was complicated and expensive. The sensor was also susceptible to magnetic and electrical interference. Therefore, there is a need to develop novel, inexpensive method for tracking the VPS in SP-based cardiac auscultations training. Using ECG signals for virtual pathology stethoscopes tracking has not been much investigated.

Most ECG related research works focus on processing the signal to monitor and diagnose cardiac abnormalities. Earlier pattern recognition methods used characteristic features extracted from the QRS and T waves [13]–[15] . However, with the advancement of computational technology, the past decade has seen the re-emergence of data-driven deep learning. Convolutional neural networks (CNNs) have proven to outperform other machine learning methods. For instance, most recent works by [16]–[18] implementing a one-dimensional CNN for ECG classification showed significant gain in detection accuracy. Hannun et al. [16] developed a deep CNN (34 layers) heart arrhythmias classifier and exceeded the performance of board-certified cardiologists.

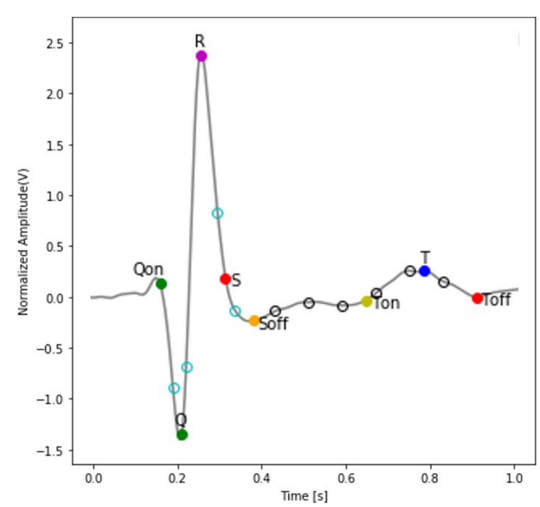

Fig. 2. Segmented ECG data from the Mitral area. Amplitude and interval features are extracted from onset, peak, and offset of QRS and T waves.

We intend to provide high realism and increase the likelihood of trainees to suspend disbelief; therefore, immediate and accurate identification of the auscultation areas is critical for fully immersive simulation experience. The ability of CNN to classifies and learn useful features from raw data without extensive feature engineering makes them well suited candidate for our application.

III. METHOD

A. Data Acquisition and Preprocessing

The hardware prototype consisted of an e-health sensor shield [19] coupled with the Raspberry Pi single board computer [20]. The data was collected using the sensor shield and an algorithm was developed on the Raspberry Pi to record and transmit ECG signal(to PC) at a sampling rate of 1KHz. Although processing and model training are done offline, the embedded platform was designed for future real-time inference were an immediate and accurate identification of the auscultation areas is critical for fully immersive simulation experience.

Noises and artifacts from breathing, body movements, and power line interference were filtered using low pass and high pass filters. A Butter-worth low pass filter with 40Hz cut-off frequency was utilized to remove the high-frequency power line noise. Similarly, a high pass FIR filter with a cut-off frequency of 0.7Hz [21] was used to reduce baseline wandering noises and low-frequency artifacts.

B. Segmentation and QRS wave Features

The Pan-Tompkins algorithm [21] was adapted to identify the well-recognized Q, R, and S peaks along with their corresponding time indices. Using a wave segmentation method [22], T wave peak was found by searching for local maxima within a pre-defined distance from the R peak. Amplitude and interval features are then extracted from the onset, peak, and offset of the QRS and T wave (Fig. 2).

These hand-crafted features may not fully capture the subtle difference between the CA areas on different subjects. In our second approach, we let a CNN automatically learn the features from the raw ECG data. Using the RR interval, we segmented the raw signal into one-second window equally centered on each R-peak. For this step, the signal was down-sampled to 200Hz (or 200 samples per second); the model was designed to predict the areas for every heartbeat.

C. Convolution Neural Network Model

We propose a neural network architecture to classify the four CA (Fig. 3). The network consists of two hierarchical 1D CNN connected in series, followed by a pooling layer and a dropout regularization. The CNN layers are grouped in two sets to give the model higher chance of learning features from the input data. A dropout layer is included to slow down the learning process for a better final model. The pooling layer reduces the learned features and consolidates them to the essential elements.Finally, the learned features are flattened to one long vector and pass through a fully connected layer for CA prediction.

The fully connected layer ideally provides a buffer between the learned features and feeds its output to the final classification layer (SoftMax Layer) that provides the four classes. Both input data (hand-crafted features or raw data) are feed to input to the network. Each cell has an inner dimension of 128 parallel feature maps and a kernel size of 5. The initial feature extraction from the input data is performed in the fully connected layer, while a second fully connected layer collects the output from the convolutional to the final classification layer.

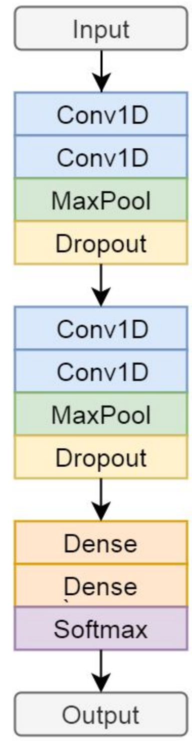

Fig. 3. The proposed model architecture of CNN

IV. EXPERIMENTAL RESULTS

A. Subjects and Acquisition Protocol

Ten standardized patients, 8 males and 2 females, from the Eastern Virginia Medical School volunteered for the study. Exclusion criteria include the presence of a pacemaker, a history of chronic or hypertension. No health-related or personal information was collected, and all subjects signed consent forms before data collection.

Data collection consists of placing the stethoscope on the bare skin over the four CA areas. The stethoscope was modified to contain two direct-contact electrode sensors fixed on to the diaphragm. The chest-piece was initially placed in a horizontal orientation and then rotated in increments of 45 degrees, producing four different orientations for each CA site; the ECG signal is symmetric, and hence, we only considered horizontal to 135°orientations. This action mimicked that of actual stethoscopy.

A total of 5 runs were collected at each auscultation site, and each run consisted of 10 seconds of ECG signals. We recorded the data from both seated and supine SP positions, giving us a total of 160 ten-second ECG recordings for each subject (4 CA areas, 4 orientations, 5 runs, and 2 postures).

B. Performance Evaluation

As stated above, we first denoise and filtered the data to preserve the ECG signal boundaries. Depending on the heart rate, each 10-second signal may contain 8 to 12 pluses. A feature vector (size = 14) is then created from the amplitude, and interval features extracted from the onset and peaks of the fiducial points (Fig. 2). We used this baseline model for comparison with our previous random forest-based setup [10]. The CNN is trained to output a class prediction for every beat. The total training dataset is divide into 80% training and 20% testing; care was taken to avoid subject overlap.

Table I reports the comparison on the model's accuracy, precision, recall (sensitivity), and F1 score. Although using a limited amount of hand-crafted features, the CNN model showed higher performance on all four statistical metrics, with an accuracy of 89% and an F1 score of 87% (a nearly 5% improvement).

TABLE I
COMPARISON OF THE CNN MODEL WITH THE PREVIOUS WORK

Classifier	Features[a]	Accuracy	Precision	Recall	F1
Random Forest [10]	QRS + T waves	0.84	0.82	0.83	0.83
CNN (current method)	QRS + T waves	0.89	0.85	0.87	0.86

[a] QRS and T- wave features.

We then trained the model with the segmented raw data, and tested the performance for seated, supine, and combined SP positions. Tables II and III depict the model performance and confusion matrix, respectively. In all CA areas, the model learned from raw data achieved higher classification accuracy compared to the model trained on hand-crafted features. For the supine position, we achieved an accuracy of 92% and an F1 score of 91%. However, the performance was lower on the seated position for all CA areas. Although efforts were made to reduce noise from the data, we observed higher irregularities (e.g. movement artifacts) on the most seated ECG data. This may contribute to the lower results, but more test needs to be done.

TABLE II
MODEL PERFORMANCE FOR DIFFERENT SP POSTURES

Data[a]	Average Model Performance(%)			
	Accuracy	Precision	Recall	F1
Seated	86.1	86.0	86.0	86.2
Supine	92.0	92.0	92.0	92.0
Combined	90.0	90.0	90.0	90.0

[a] 1 sec row data (sampled at 200 pts/sec) segmented between each RR interval.

V. CONCLUSION

We investigated the accuracy and validity of a CNN based virtual pathology stethoscope tracking system. This involved collecting and examining ECG data from multiple subjects. The preliminary analysis done on 10 subjects, showed promising performance gain over our previous system.

TABLE III
CONFUSION MATRIX

Aortic	495	10	6	12
Mitral	6	515	9	4
Pulmonic	13	5	503	12
Tricuspid	7	5	8	551

However, testing on additional subjects as well as different postures is required to determine the reliability of the system. The information gained from this study would guide us in re-designing the embedded platform for real-time inference were an immediate and accurate identification of the auscultation areas is critical for fully immersive simulation experience.

REFERENCES

[1] P. A. Heidenreich et al., "Forecasting the Future of Cardiovascular Disease in the United States A Policy Statement From the American Heart Association," (in English), Circulation, vol. 123, no. 8, pp. 933-944, Mar 1 2011.

[2] A. Minino, E. Arias, K. Kochanek, S. Murphy, and B. Smith, "Deaths: final data for 2000. National vital statistics reports," Hyattsville, MD: National Center for Health Statistics, 2002.

[3] R. A. Orourke, "Cardiac Auscultation - a Cost-Effective Diagnostic Skill," Current Problems in Cardiology, vol. 20, no. 7, pp. 445-530, Jul 1995.

[4] R. J. Ostfeld, Y. H. Goldberg, G. Janis, S. Bobra, H. Polotsky, and S. Silbiger, "Cardiac auscultatory training among third year medical students during their medicine clerkship," International journal of cardiology, vol. 144, no. 1, pp. 147-149, 2010.

[5] S. Mangione and L. Z. Nieman, "Cardiac auscultatory skills of internal medicine and family practice trainees: a comparison of diagnostic proficiency," Jama, vol. 278, no. 9, pp. 717-722, 1997.

[6] J. M. Vukanovic-Criley et al., "Competency in cardiac examination skills in medical students, trainees, physicians, and faculty: a multicenter study," Archives of internal medicine, vol. 166, no. 6, pp. 610-616, 2006.

[7] N. Kidane, S. Chemlal, J. Li, F. D. McKenzie, and T. Hubbard, "Orientation invariant ECG-based stethoscope tracking for heart auscultation training on augmented standardized patients," Simulation, vol. 89, no. 12, pp. 1450-1458, 2013.

[8] V. L. Clark and J. A. Kruse, "Clinical methods: the history, physical, and laboratory examinations," Jama, vol. 264, no. 21, pp. 2808-2809, 1990.

[9] J. J. Ward and B. A. Wattier, "Technology for enhancing chest auscultation in clinical simulation," Respiratory care, vol. 56, no. 6, pp. 834-845, 2011.

[10] A. Cross, N. Kidane, E. M. Leviste, T. Hubbard, and F. McKenzie, "Board 322-Research Abstract Validation of an ECG-Based Stethoscope Tracking Technology (Submission# 1247)," Simulation in Healthcare, vol. 8, no. 6, pp. 528-825, 2013.

[11] A. Castilano, N. Haller, C. Goliath, and P. Lecat, "The VentriloscopeÂ®: âAm I hearing things?â," Medical Teacher, vol. 31, no. 3, pp. e97-e101, 2009.

[12] R. Castelino, "Augmented Reality for Auscultation in Standardized Patients," Master, Electrical Engineering, OLD DOMININON UNIVERSITY, USA, 2005.

[13] T. T. Khan, N. Sultana, R. B. Reza, and R. Mostafa, "ECG feature extraction in temporal domain and detection of various heart conditions," in 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2015, pp. 1-6: IEEE.

[14] K.-i. Minami, H. Nakajima, and T. Toyoshima, "Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network," IEEE transactions on Biomedical Engineering, vol. 46, no. 2, pp. 179-185, 1999.

[15] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," IEEE transactions on biomedical engineering, vol. 51, no. 7, pp. 1196-1206, 2004.
[16] A. Y. Hannun et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," Nat Med, vol. 25, no. 1, pp. 65-69, Jan 2019.
[17] S. Kiranyaz, T. Ince, , and M. Gabbouj, âReal-time patient-specific ecg classification by 1-d convolutional neural networks,â IEEE Transactions on Biomedical Engineering, vol. 63, no. 3, pp. 664â675, March 2016.
[18] S. Saadatnejad, M. Oveisi, and M. Hashemi, "LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices," IEEE journal of biomedical and health informatics, 2019.
[19] A. "e-health sensor platform V2. 0 for Arduino and Raspberry Pi," ed: cooking-hacks, 2015.
[20] A. "Raspberry Pi Model B," ed: ed, 2015.
[21] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," IEEE Trans. Biomed. Eng, vol. 32, no. 3, pp. 230-236, 1985.
[22] A. T. T. Khan, N. Sultana, R. B. Reza, and R. Mostafa, "ECG feature extraction in temporal domain and detection of various heart conditions," in 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2015, pp. 1-6: IEEE.