

ECG-BASED VIRTUAL PATHOLOGY STETHOSCOPE TRACKING USING TRANSFER LEARNING

Haben Yhdego
Nahom Kidane
Frederick Mckenzie
Michel Audette

Old Dominion University
5115 Hampton Boulevard
Norfolk, VA, USA
{hyhde001,nkidane,rdmckenz,maudette}@odu.edu

ABSTRACT

Standardized Patient (SP) based medical simulation is commonly used to teach bedside skills. However, SPs are typically healthy individuals, and the number and range of conditions they can portray are limited. Our research aims to improve the cardiac auscultation (CA) skills of medical students by utilizing a modified stethoscope in tandem with SP. This technology introduces the potential to augment virtual pathology sounds on otherwise healthy SP. In this study, CNN models, previously trained on large-scale image datasets, are transferred to identify the four CA sites. We applied the pre-trained CNN models of ResNet50, InceptionV3, and Alexnet, to the ECG recordings from the CA regions, which are converted to images using a wavelet scalogram. Moreover, data augmentation is performed to supplement limited labeled data. Experimental results illustrate data augmentation with InceptionV3 and Resnet50 models leads to a better performance than our previously reported methods, between 93% and 95% F1 score.

Keywords: ECG, Transfer Learning, Clinical Simulation, Auscultation, Stethoscope Tracking

1 INTRODUCTION

Cardiac auscultation (CA) is the auditory detection of heart sounds (Figure 1) to diagnose abnormalities, a crucial skill that is both efficient and cost-effective in medical practice. Cardiac diseases such as congestive heart failure, systemic arterial hypertension, and coronary artery disease can be diagnosed with proper CA (Vukanovic-Criley et al. 2006). However, several surveys (Vukanovic-Criley et al. 2006), (Chizner 2008), (Mangione 2001), (Spatz et al. 2011), (Wayne et al. 2009)), have highlighted a rapid decline in this CA skill among medical trainees and residents. A study ((Vukanovic-Criley et al. 2006)) conducted in 2006, showed that CA skills among trainees did not improve after the 3rd year of medical studies. Wayne and his colleagues surveyed 100 medical students, found that only 1/3 of the trainees passed the minimum passing standard in CA on the first try (Wayne et al. 2009) . The main reasons cited for this decline is limited patient access due to health-care management restrictions, shortage in training time, and reliance in competing & more expensive technologies.



Figure 1: Auscultation.

Medical simulators using Standardized Patients (SPs) have been used to teach CA, which provides a safe and controlled environment for repetitive practice and immediate training opportunities. However, SPs are typically healthy individuals, and hence the number and range of conditions they can portray are limited. Our research aims to improve the auscultation skills of medical students by virtual pathology simulation that utilizes modified stethoscopes in tandem with standardized patients. This technology introduces the potential to augment virtual pathology (abnormal sounds) on otherwise healthy SP.

Currently, an electrocardiogram(ECG)-based tracking prototype stethoscope is being tested to simulate a realistic cardiac examination (Kidane et al. 2013). It takes advantage of subtle ECG signal differences among the four main CA sites (Aortic, Pulmonic, Tricuspid, and Mitral areas) to identify which spot the stethoscope head is placed upon. The system recognizes these locations using different machine learning algorithms (Figure 2) . But, the current results do not provide a realistic experience for medical students as the highest accuracy achieved using a random forest classifier was 86%.

In this work, CNN models, previously trained on large-scale general image datasets, are transferred to identify the four CA sites. We applied the pre-trained CNN models of Alexnet (Krizhevsky et al. 2012), ResNet50 (He et al. 2015), and InceptionV3 (Szegedy et al. 2015) to the ECG datasets, which are converted to RGB images using a time-frequency wavelet scalogram. Moreover, data augmentation is performed on the scalogram images to supplement the limited labeled data. This paper is organized as follows: Section II will present the methods for comparison and data collection. Section III will highlight experimental results and subsequent model selection for the relevant work. Finally, we will discuss conclusory remarks and propose potential future work.

2 METHODS

One of the main goals of the proposed application is to make real time augmented reality application that can run on top of relatively cheap and resource-constrained devices, such as a micro-controller equipped smart sensor, an embedded platform or a wearable sensors. Therefore, the selection of the deep learning architecture that implements the inference module, i.e. the component that performs the actual detection, must take into account this requirement. The overall procedures of the proposed system is shown in Figure 2.

2.1 Data Acquisition

A modified stethoscope apparatus recorded ECG signals from the four auscultation sites of an SPs chest. The setup consisted of a double-lead configuration and an additional base lead that was placed on the Sps leg. Data was collected using an e-health sensor shield attached to a Raspberry Pi single-board computer.

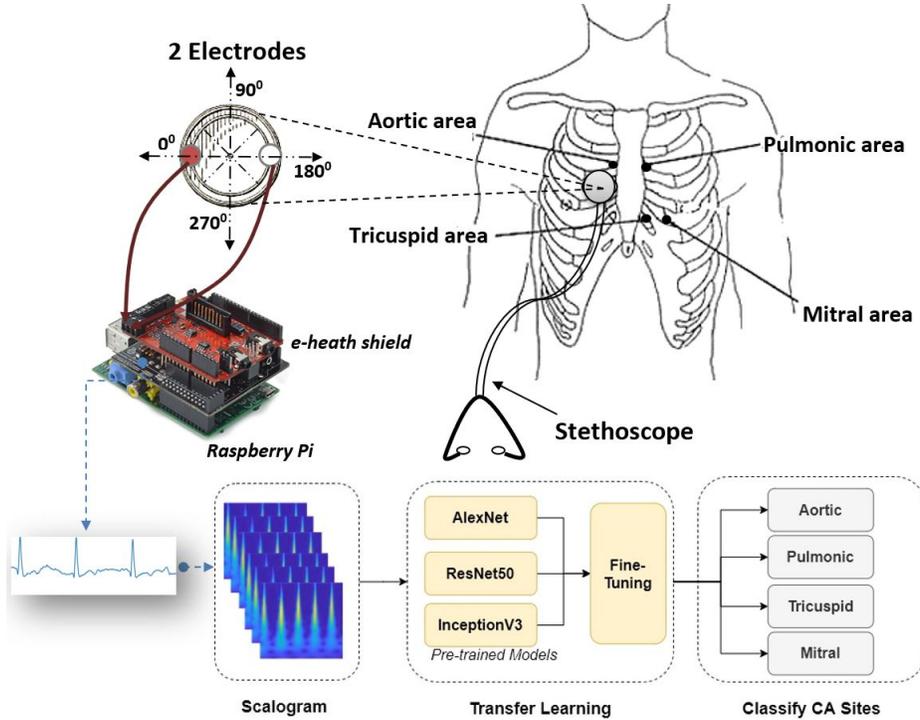


Figure 2: Proposed System.

The embedded platform was designed for future real-time inference where an immediate identification is critical for a fully immersive simulation. Ten healthy standardized patients, 8 males and 2 females, from the Eastern Virginia Medical School, volunteered for the study. To test for stethoscope orientation, signals were collected using angular variations but rotating the stethoscope head clockwise with increments of 45° to produce a total of 4 orientations per CA site. A total of 5 runs of ECG data were collected per angle at each CA site, and each run consisted of 10 beats of signals. Data was collected for both supine and seated positions, giving us a total of 160 ten-second ECG recordings for each subject (4 CA areas, 4 orientations, 5 runs, and 2 postures).

2.2 Continuous Wavelet Transform

Previous works have combined continuous wavelet transform (CWT) representation with CNN to address classification tasks of time-series data (Ravi et al. 2016) (Yhdego et al. 2019). When calculating the scalogram of the ECG signal, the information can be structured as a $\text{Time} \times \text{Spatial} \times \text{Scale}$. Mathematically, CWT represents time-varying spectral information of a continuous signal by computing the inner products of this signal with a set of continuous wavelets. We pre-compute the CWT filter bank to create time-frequency representations of the ECG signal. This representation is then aggregated into a scalogram, which is essential for extracting interpretable features that capture the intensity differences among the nearest data points. It is also time and sampling-rate invariant, enabling the system to be more robust against data shifts. Moreover, frequency selection in the scalogram domain also provides an implicit noise filtering capability.

To use transfer learning on the ECG datasets, we transform the 1D signal data to RGB images. This conversion is necessary because our transfer learning approach exploits the feature vectors from the pre-trained model of the ImageNet datasets (Russakovsky et al. 2015) and the different architectures that are trained to

classify images. We reuse three different network architecture of the CNN to classify ECG signals based on scalogram images from the CWT of the time series data. Our inputs to the model architectures are the scalograms RGB images. Before the data were inputted into the network, each image is transformed to an array of size $299 \times 299 \times 3$ for inceptionV3 and $224 \times 224 \times 3$ for ResNet50 and alexnet.

2.3 Transfer Learning with CNN

CNN is difficult to apply in the medical field since collecting enough extensive data from patients is challenging. However, transfer learning can be applied in such cases where there is limited training data. For image recognition, the early layers of the network typically learn low-level features such as edges and corners, and further layers learn high-level features such as textures and objects ((Zeiler and Fergus 2013)). Hence, the pre-trained initial layers can be used to extract low-level features, and the last few layers can be retrained and tailored to extract specific features related to our limited ECG dataset.

Three different state of the art CNN architectures were implemented in TensorFlow: Alexnet (Krizhevsky et al. 2012), ResNet50 (He et al. 2015), and InceptionV3 (Szegedy et al. 2015). They were all pre-trained on the ImageNet dataset (Russakovsky et al. 2015).

Alexnet: AlexNet deep CNN contains 5 convolutional layers, 3 fully connected layers, and dropout applied before the first and second fully connected layer. The network achieved the state of the art performance (15.3% VS 26.2% for the second place) in the 2012 ImageNet ILSVRC-2012 competition.

Resnet50: The Resnet has 49 convolutional layers and one fully connected layer (He et al. 2015). It uses residual mapping which aid in avoiding degradation problems that occurs with very deep CNNs. In addition, the intermediate normalization layers of the Resnet also solves the problem of vanishing and exploding gradients. The reader is referred to article (He et al. 2015) for detailed explanation.

InceptionV3: The Inception network used in this work is from InceptionV3 (Szegedy et al. 2015), which had 95 layers. The architecture is designed to overcome the difficulty of choosing the filter size or when to use the pooling layer. It solves this by using different filter sizes and pooling layers in parallel. Computational complexity is reduced by using a one by one convolutions to shrink the volume of the next layer. The network was introduced by Szegedy and his colleagues (Szegedy et al. 2015).

Fine-tuning the Pre-trained Model

As described above, instead of random parameter initialization, our networks were all pre-trained on the ImageNet dataset. We then fine-tuned the networks to fit the ECG datasets better. The last three layers of the three architectures discussed above are changed and retrained for classifying the RGB image of our ECG dataset. The fully connected layer, initially configured for 1000 categories, is set to four classes representing the CA areas.

3 EXPERIMENT RESULTS

Data was obtained from 10 recruited Standardized Patients (SP) in Eastern Virginia Medical School's Sentara Center for Simulation and Immersive Learning to conduct 2 in-person trials of the ECG-based tracking stethoscope (Cross et al. 2013). The participants involved were not required to complete any prior training or evaluations of the appointment, nor did they act out any symptoms of cardiac abnormalities during data acquisition. The use of SPs for this study was strictly for obtaining ECG signal data and testing the accuracy and real-time capabilities of site detection.

Furthermore, a data augmentation procedure involving random rotations was used. It introduces a label-invariant variability of wearable sensor data (Taewoong et al. 2017). For example, downside placement of the sensor can invert the sign of the sensor readings without changing the labels. Therefore, augmentation by applying arbitrary rotations to the existing data can be used as a way of simulating different sensor placements. Given the sensor ECG signal data vector at time t , a new vector can be obtained by rotating by 5 degrees (Yhdego et al. 2019).

Table 1: Comparing the results of different methods.

	Precision	Recall	F1-Score
Inception V3 + DataAug	0.96	0.94	0.95
Resnet50 + DataAug	0.94	0.92	0.93
Alexnet + DataAug	0.92	0.90	0.91
Inception V3	0.87	0.85	0.86
Resnet50	0.88	0.86	0.87
Alexnet	0.90	0.88	0.89

In order to evaluate the performance of the classifiers, two experiments were conducted. The first evaluates the models using the ECG datasets with data augmentation, whereas the second aims at using the ECG datasets only. As shown in Table 1, three of the models that use the data augmentation outperform those that do not use data augmentation. The F1-score of InceptionV3 and Resent50 without data augmentation is 87% and 86%, whereas, with data augmentation, it increases to 93% and 95%, respectively. In the case of Alexnet architecture, there is no huge difference with or without data augmentation, which is 91% and 89%, respectively. Because the InceptionV3 and Resent50 networks are deeper than the Alexnet network, it needs more data to get better performance.

We illustrate the accuracy of these predictions by plotting confusion matrices for all models, as shown in Figure 3. InceptionV3, ResNet50, and Alexnet with data augmentation attained mean accuracy of 94%, 92%, and 90%, respectively. Thus comparing with previous algorithms that use Random forest (Kidane et al. 2013), 86% was obtained for the four auscultation sites in terms of accuracy, where our ResNet50 and InceptionV3 applications reached greater accuracy than any model yet reported.

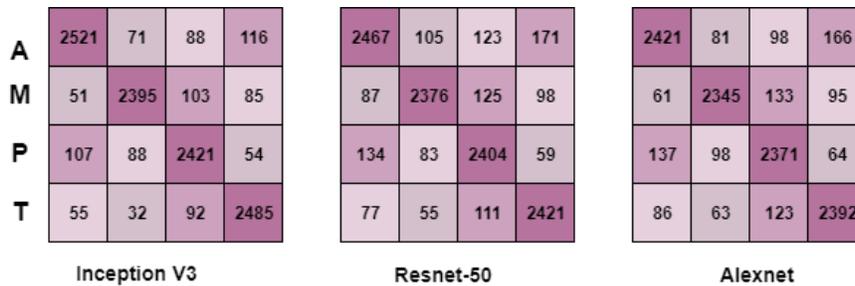


Figure 3: Confusion matrices for the different models.

4 CONCLUSION

In this paper, we use transfer learning methods to boost the performance of an orientation invariant ECG-based virtual pathology stethoscope tracking method for placing virtual symptoms in correct auscultation landmarks of patient. The experimental results demonstrate that inceptionV3 model leads to a better performance than the existing methods in the case of scarce labeled training data and the additional data augmentation helps to boost the capability of classification results. Furthermore, data augmentation by applying arbitrary rotations to the existing ECG data helps as a way of simulating different sensor placements.

In our future work, the classification results that we get from the supine dataset can be further compared with the seated dataset and combined both postures to determine the reliability of the method. Sequentially classifying the signals must also be studied; such a system will perform real-time classification and make predictions on the incoming beat or section of an ECG signal. For example, if the current signal is from a particular region, the next signal has a higher likelihood of being from the same region.

REFERENCES

- Chizner, M. A. 2008. "Cardiac auscultation: rediscovering the lost art". *Curr Probl Cardiol* vol. 33 (7), pp. 326–408.
- Cross, A., N. Kidane, E. Leviste, T. Hubbard, and F. McKenzie. 2013. *Validation of an ECG-Based Stethoscope Tracking Technology*.
- He, K., X. Zhang, S. Ren, and J. Sun. 2015. "Deep Residual Learning for Image Recognition". *CoRR* vol. abs/1512.03385.
- Kidane, N., S. Chemlal, J. Li, F. D. McKenzie, and T. Hubbard. 2013. "Orientation invariant ECG-based stethoscope tracking for heart auscultation training on augmented standardized patients". *SIMULATION* vol. 89 (12), pp. 1450–1458.
- Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks". *26th Annual Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, United States*, pp. 1106–1114.
- Mangione, S. 2001. "Cardiac auscultatory skills of physicians-in-training: a comparison of three English-speaking countries". *Am J Med* vol. 110 (3), pp. 210–6.
- Ravi, D., C. Wong, B. Lo, and G. Yang. 2016. "Deep learning for human activity recognition: A resource efficient implementation on low-power devices". in *Proc. 2016 IEEE 13th Int. Conf. Wearable Implantable Body Sensor Netw.*, pp. 71–76.
- Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. 2015. "ImageNet Large Scale Visual Recognition Challenge". *International Journal of Computer Vision* vol. 115 (3), pp. 211–252.
- Spatz, E. S., D. LeFrancois, and R. J. Ostfeld. 2011. "Developing cardiac auscultation skills among physician trainees". *Int J Cardiol* vol. 152 (3), pp. 391–2.
- Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. 2015. "Rethinking the Inception Architecture for Computer Vision". *CoRR* vol. abs/1512.00567.
- Taewoong, T., F. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. Fietzek, and D. Kulic. 2017. "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks". *ICMI*, DOI:10.1145/3136755.3136817.
- Vukanovic-Criley, J., S. Criley, C. Warde, J. Boker, L. Guevara-Matheus, W. Churchill, W. Nelson, and J. Criley. 2006. "Competency in cardiac examination skills in medical students, trainees, physicians, and faculty: a multicenter study". *Arch Intern Med* vol. 166 (6), pp. 610–6.
- Wayne, D. B., J. Butter, E. R. Cohen, and W. C. McGaghie. 2009. "Setting defensible standards for cardiac auscultation skills in medical students". *Acad Med* vol. 84 (10 Suppl), pp. S94–6.
- Yhdego, H., J. Li, S. Morrison, M. Audette, C. Paolini, M. Sarkar, and H. Okhravi. 2019. "Towards Musculoskeletal Simulation-Aware Fall Injury Mitigation: Transfer Learning with Deep CNN for Fall Detection". In *2019 Spring Simulation Conference (SpringSim)*, pp. 1–12.
- Zeiler, M. D., and R. Fergus. 2013. "Visualizing and understanding convolutional networks.". *CoRR* abs/1311.2901.

AUTHOR BIOGRAPHIES

HABEN YHDEGO is a PhD student in the Computational Modeling and Simulation Engineering (CMSE) Department at Old Dominion University. He received MSc in Image processing and computer vision from Jean-Monet University, Saint-Etienne, France. His research interests includes applying machine learning methods for images and inertia sensor datasets, and computer vision. His email address is hyhde001@odu.edu.

NAHOM KIDANE is a Ph.D. candidate and an Adjunct Professor at Old Dominion University's Computational Modeling and Simulation Engineering Department. He received his BS degree in Electrical Engineering from Jimma University, Jimma, Ethiopia and his MS degree in Modeling and Simulation Engineering from Old Dominion University, Norfolk, Virginia. His main areas of research are medical simulation, 3D deep learning, augmented reality, and scientific visualization. His email address is nkida001@odu.edu.

RICK MCKENZIE was born in Kingston, Jamaica. Dr. McKenzie has a B.S. in engineering, M.S. in computer engineering, and a Ph.D. in computer engineering all from the University of Central Florida in Orlando, Florida U.S.A. Currently, he is the Department Chair in the Computational Modeling and Simulation Engineering (CMSE) Dept. and a joint faculty member in the Department of Electrical and Computer Engineering (ECE) at Old Dominion University in Norfolk, Virginia U.S.A. In addition, Dr. McKenzie is the co-director of the Medical Imaging, Diagnosis, and Analysis (MIDA) Laboratory, Adjunct Associate Professor of Eastern Virginia Medical School (EVMS) in the School of Health Professions, and Graduate Faculty Scholar at University of Central Florida (UCF) in the School of Electrical Engineering and Computer Science. Before coming to Old Dominion University, he spent 6 years in the simulation industry at Science Applications International Corporation (SAIC) as a senior scientist. Dr. McKenzie's research has been in medical modeling and simulation, human behavior representation, and simulation architectures often focusing on aspects of scientific visualization and virtual reality. His email address is rdmckenz@odu.edu.

MICHEL AUDETTE is an Associate professor of the Department of Computational Modeling and Simulation Engineering (CMSE) at Old Dominion University. He holds a PhD in Biomedical Engineering from McGill University. His research interests include medical simulation, surgery planning and medical image analysis. His email address is maudette@odu.edu.