when compared with features gained from the pre-trained VGG16 model.

4 CONCLUSIONS

We proposed to apply and adapt pre-trained *Image Classification Convolutional Neural Networks* (ImageNet) on scalogram images of *Phonocardiogram* (PCG) for the task of normal/ abnormal heart sound classification. Deep PCG representations extracted from a task-adapted version of the popular ImageNet VGG16 were shown to be more robust for this task than the widely used ComParE audio feature set. The combination of learnt VGG features and a *Support Vector Machine* (SVM) significantly ($p < .001$ by one-tailed z-test) outperformed the ComParE based baseline system. We speculate this success is due to the autonomous nature of the feature extraction associated with the 'learnt VGG' topology; the representations are adapted to the dataset and therefore are more robust than a 'fixed' conventional feature set.

In future work, data augmentation will be investigated for heart sound classification to compensate for the unbalanced nature of the dataset. Further, a new ImageNet topology based on the scalogram images will be developed and validated on a variety of heart sound datasets, e. g., AudioSet[4], to build a robust ImageNet for heart sound classification.

ACKNOWLEDGMENTS


This work was partially supported by the German national BMBF IKT2020-Grant under grant agreement No. 16SV7213 (EmotAsS), the EU Horizon 2020 Innovation Action Automatic Sentiment Estimation in the Wild under grant agreement No. 645094 (SEWA), and the EU H2020/ EFPIA Innovative Medicines Initiative under grant agreement No. 115902 (RADAR-CNS).

REFERENCES

[1] Shahin Amiriparian, Maurice Gerczuk, Sandra Ottl, Nicholas Cummins, Michael Freitag, Sergey Pugachevskiy, Alice Baird, and Björn Schuller. 2017. Snore sound classification using image-based deep spectrum features. In *Proc. INTERSPEECH*. Stockholm, Sweden, 3512–3516.

[2] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2, 3 (Apr. 2011), 1–27.

[3] Gari D. Clifford, Chengyu Liu, Benjamin Moody, David Springer, Ikaro Silva, Qiao Li, and Roger G. Mark. 2016. Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016. In *Proc. Computing in Cardiology Conference (CinC)*. Vancouver, Canada, 609–612.

[4] Jun Deng, Nicholas Cummins, Jing Han, Xinzhou Xu, Zhao Ren, Vedhas Pandit, Zixing Zhang, and Björn Schuller. 2016. The University of Passau open emotion recognition system for the multimodal emotion challenge. In *Proc. CCPR*. Chengdu, China, 652–666.

[5] Florian Eyben, Felix Weninger, Florian Groß, and Björn Schuller. 2013. Recent Developments in openSMILE, the Munich open-source multimedia feature extractor. In *Proc. ACM Multimedia*. Barcelona, Spain, 835–838.

[6] Steve R. Gunn. 1998. Support vector machines for classification and regression. *ISIS technical report* 14, 1 (May 1998), 5–16.

[7] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. 2014. Large-scale video classification with convolutional neural networks. In *Proc. CVPR*. Columbus, OH, 1725–1732.

[8] Edmund Kay and Anurag Agarwal. 2016. DropConnected neural network trained with diverse features for classifying heart sounds. In *Proc. Computing in Cardiology Conference (CinC)*. Vancouver, Canada, 617–620.

[9] Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882* (2014).

[4] https://research.google.com/audioset/dataset/heart_sounds_heartbeat.html

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Proc. NIPS*. Lake Tahoe, NV, 1097–1105.

[11] Aubrey Leatham. 1952. Phonocardiography. *British Medical Bulletin* 8, 4 (1952), 333–342.

[12] Chengyu Liu et al. 2016. An open access database for the evaluation of heart sound algorithms. *Physiological Measurement* 37, 12 (Nov. 2016), 2181–2213.

[13] Ilias Maglogiannis, Euripidis Loukis, Elias Zafiropoulos, and Antonis Stasis. 2009. Support vectors machine-based identification of heart valve diseases using heart sounds. *Computer Methods and Programs in Biomedicine* 95, 1 (July 2009), 47–61.

[14] Vykintas Maknickas and Algirdas Maknickas. 2017. Recognition of normal–abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients. *Physiological Measurement* 38, 8 (July 2017), 1671–1679.

[15] Ali Moukadem, Alain Dieterlen, Nicolas Hueber, and Christian Brandt. 2013. A robust heart sounds segmentation module based on S-transform. *Biomedical Signal Processing and Control* 8, 3 (May 2013), 273–281.

[16] Dariush Mozaffarian et al. 2016. Heart disease and stroke statistics–2016 update: A report from the American Heart Association. *Circulation* 133, 4 (Jan. 2016), e38–e360.

[17] Sofia C. Olhede and Andrew T. Walden. 2002. Generalized morse wavelets. *IEEE Transactions on Signal Processing* 50, 11 (Nov. 2002), 2661–2670.

[18] Chrysa D. Papadaniil and Leontios J. Hadjileontiadis. 2014. Efficient heart sound segmentation and extraction using ensemble empirical mode decomposition and kurtosis features. *IEEE Journal of Biomedical and Health Informatics* 18, 4 (July 2014), 1138–1152.

[19] Kun Qian, Christoph Janott, Vedhas Pandit, Zixing Zhang, Clemens Heiser, Winfried Hohenhorst, Michael Herzog, Werner Hemmert, and Björn Schuller. 2017. Classification of the excitation location of snore sounds in the upper airway by acoustic multifeature analysis. *IEEE Transactions on Biomedical Engineering* 64, 8 (Aug. 2017), 1731–1741.

[20] Kun Qian, Christoph Janott, Zixing Zhang, Clemens Heiser, and Björn Schuller. 2016. Wavelet features for classification of vote snore sounds. In *Proc. ICASSP*. Shanghai, China, 221–225.

[21] Zhao Ren, Vedhas Pandit, Kun Qian, Zijiang Yang, Zixing Zhang, and Björn Schuller. 2017. Deep sequential image features on acoustic scene classification. In *Proc. DCASE Workshop*. Munich, Germany, 113–117.

[22] Olivier Rioul and Martin Vetterli. 1991. Wavelets and signal processing. *IEEE Signal Processing Magazine* 8, 4 (Oct. 1991), 14–38.

[23] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* 115, 3 (Dec. 2015), 211–252.

[24] Maryam Samieinasab and Reza Sameni. 2015. Fetal phonocardiogram extraction using single channel blind source separation. In *Proc. ICEE*. Tehran, Iran, 78–83.

[25] Samuel E. Schmidt, Claus Holst-Hansen, Claus Graff, Egon Toft, and Johannes J. Struijk. 2010. Segmentation of heart sound recordings by a duration-dependent hidden Markov model. *Physiological Measurement* 31, 4 (Mar. 2010), 513–529.

[26] Samuel E. Schmidt, Claus Holst-Hansen, John Hansen, Egon Toft, and Johannes J. Struijk. 2015. Acoustic features for the identification of coronary artery disease. *IEEE Transactions on Biomedical Engineering* 62, 11 (Nov. 2015), 2611–2619.

[27] Samuel E. Schmidt, Egon Toft, Claus Holst-Hansen, and Johannes J. Struijk. 2010. Noise and the detection of coronary artery disease with an electronic stethoscope. In *Proc. CIBEC*. Cairo, Egypt, 53–56.

[28] Björn Schuller, Stefan Steidl, Anton Batliner, Elika Bergelson, Jarek Krajewski, Christoph Janott, Andrei Amatuni, Marisa Casillas, Amdanda Seidl, Melanie Soderstrom, Anne Ss Warlaumont, Guillermo Hidalgo, Sebastian Schnieder, Clemens Heiser, Winfried Hohenhorst, Michael Herzog, Maximilian Schmitt, Kun Qian, Yue Zhang, George Trigeorgis, Panagiotis Tzirakis, and Stefanos Zafeiriou. 2017. The INTERSPEECH 2017 computational paralinguistics challenge: Addressee, cold & snoring. In *Proc. INTERSPEECH*. Stockholm, Sweden, 3442–3446.

[29] Björn Schuller, Stefan Steidl, Anton Batliner, Alessandro Vinciarelli, Klaus Scherer, Fabien Ringeval, Mohamed Chetouani, Felix Weninger, Florian Eyben, Erik Marchi, Marcello Mortillaro, Hugues Salamin, Anna Polychroniou, Fabio Valente, and Samuel Kim. 2013. The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism. In *Proc. INTERSPEECH*. Lyon, France, 148–152.

[30] Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. In *Proc. ICLR*. San Diego, CA, no pagination.

[31] Zeeshan Syed, Daniel Leeds, Dorothy Curtis, Francesca Nesta, Robert A. Levine, and John Guttag. 2007. A framework for the analysis of acoustical cardiac signals. *IEEE Transactions on Biomedical Engineering* 54, 4 (Apr. 2007), 651–662.

[32] Zeeshan Hassan Syed. 2003. *MIT automated auscultation system.* Ph.D. Dissertation. Massachusetts Institute of Technology.

[33] Hong Tang, Ting Li, Tianshuang Qiu, and Yongwan Park. 2012. Segmentation of heart sounds based on dynamic clustering. *Biomedical Signal Processing and Control* 7, 5 (Sep. 2012), 509–516.