On-device artificial intelligence: mobile solution for detecting severe aortic valve stenosis based on heart sounds

H. Makimoto¹, T. Shiraga², B. Kohlmann¹, C.-E. Magnisali¹, R. Schenk¹, S. Gerguri¹, N. Motoyama², L. Clasen¹, A. Bejinariu¹, J. Schmidt¹, C. Brinkmeyer¹, R. Westenfeld¹, T. Zeus¹, M. Kelm³

¹Heinrich Heine University, Medical Faculty, Division of Cardiology, Pulmonology and Vascular Medicine, Dusseldorf, Germany; ²Mitsubishi Electric Inc., Kamakura, Japan; ³Heinrich Heine University, Medical Faculty, Division of Cardiology, Pulmonology and Vascular Medicine, CARID, Dusseldorf, Germany

Funding Acknowledgement: Type of funding sources: None.

Background: Aortic stenosis is still one of the major causes of sudden cardiac death in the elderly. Noninvasive screening for severe aortic valve stenosis (AS) may result in early cardiac diagnostic leading to an appropriate and timely medical intervention.

Purpose: The aims of this study were 1) to develop an artificial intelligence to detect severe AS based on heart sounds and 2) to build an application to screen patients using electronic stethoscope and smartphones, which will provide an efficient diagnostic workflow for screening as a complementary tool in daily clinical practice.

Methods: We enrolled 100 patients diagnosed with severe AS and 200 patients without severe AS (no echocardiographic sign of AS [n=100], mild AS [n=50], moderate AS [n=50]). The heart sounds were recorded in 4000 Hz waveform audio format at the following 3 sites of each patient: the 2nd intercostal right sternal border, the Erb’s area and the apex. Each record was divided into multiple data of 4 seconds duration, which built 10800 sound records in total. We developed multiple convolutional neural networks (CNN) designed to recognize severe AS in heart sounds according to the recorded 3 sites. We adopted a stratified 4-fold cross-validation method by which the CNN was trained with 60% of the whole data, validated with 20% data and tested with the remaining 20% data not used during training and validation. As performance metrics we adopted the accuracy, F1 value and the area under the curve (AUC) calculated as the average of all cross-validation folds.

For the smartphone application, we combined the best CNN-models from each recorded site for the best performance. Further 40 patients were newly enrolled for its clinical validation (no AS [n=10], mild AS [n=10], moderate AS [n=10], severe AS [n=10]).

Results: The accuracy, F1 value and AUC of each model were 88.9±5.7%, 0.888±0.006 and 0.953±0.008, respectively. The sensitivity and specificity were 87.9±2.2% and 89.9±2.4%. The recognition accuracy of moderate AS was significantly lower as compared to the other AS grades (moderate AS 74.1±6.1% vs no AS 98.0±1.4%, mild AS 97.6±1.2%, severe AS 87.9±2.2%, respectively, P<0.05).

Our smartphone application showed a sensitivity of 100% (10/10), a specificity of 73.3% (22/30), and an accuracy of 80.0% (32/40), which implicated a good utility for screening. In the detailed analysis of 8 mistaken decisions, these were highly affected by the presence of severe mitral or tricuspid valve regurgitation despite of non-severe AS (7/8 [87.5%]).

Conclusions: This study demonstrated the promising possibility of an end-to-end screening for severe aortic valve stenosis using an electronic stethoscope and a smartphone application. This technology may improve the efficacy of daily medicine particularly where the human resource is limited or support a remote medical consultation. Further investigations are necessary to increase accuracy.