Deep Learning for ECG Screening of Moderate to Severe Tricuspid Valve Regurgitation

Yotam Hadari, Michal Cohen-Shelly, Karin Sudri, Nitsan Halabi, Miriam Gibli, Sagit Benzekry, Rafael Kuperstein, Aias Masalha, Robert Klempfner, Avi Sabbag

Aim and Background

Mitral regurgitation (MR) is a common valvular heart disease requiring timely detection to optimize treatment and prevent progression to severe complications. Early intervention, particularly in asymptomatic patients with moderate MR, can reduce long-term morbidity. Electrocardiograms (ECGs), given their accessibility and cost-effectiveness, offer a promising platform for applying artificial intelligence (AI) to enhance MR detection.

Methods

Between 2009 and 2024, we identified 49,254 adults [mean age 67 ± 16.2 years; 19,929 women (39%)] who underwent both echocardiography and ECG within a 14-day interval at Sheba Medical Center. Moderate to severe MR was identified in 3,862 Echo-ECG pairs (7.1%). A Convolutional Neural Network (CNN) was trained on 29,552 samples (60%), validated on 9,851 samples (20%), and tested on 9,851 samples (20%), stratified by MR severity, age, and sex. Data augmentation techniques and a Multi-Layer Perceptron (MLP) module were incorporated to improve performance. Additionally, mild and mild-to-moderate MR cases were excluded from the test data to refine the model's predictive accuracy.

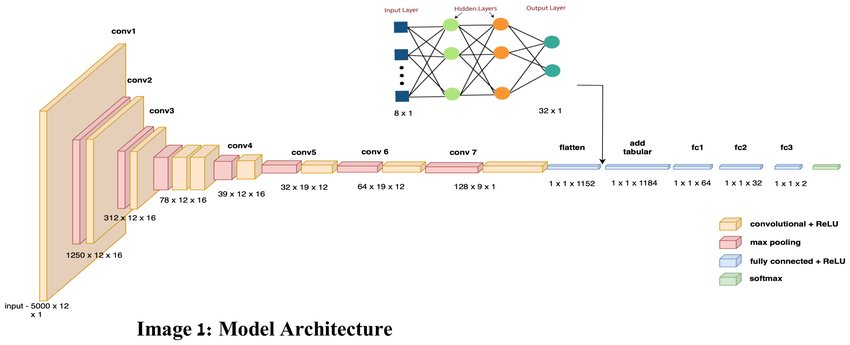

Image 1: Model Architecture

Results

The CNN achieved an AUC of 0.77, with sensitivity, specificity, and accuracy of 76.8%, 66.6%, and 67%, respectively. Incorporating data augmentation improved the AUC to 0.79, with corresponding sensitivity, specificity, and accuracy of 71.1%, 72.9%, and 72%. Adding the MLP module and excluding mild MR cases further enhanced performance, achieving an AUC of 0.80, sensitivity of 73.4%, specificity of 74.1%, and accuracy of 73%. Subgroup analysis revealed higher sensitivity in males (77.7%) compared to females (72.4%) and the strongest diagnostic performance in patients aged 40–59 years (AUC 0.818), while the lowest was observed in those under 40 years (AUC 0.734).

Model	AUC	Accuracy	Sensitivity	Specificity
CNN	0.77	67%	76.7%	66.6%
CNN + Augmentations	0.79	72%	71.1%	72.9%
CNN + augmentations +MLP + test exclusion	0.8	73%	73.4%	74.1%

Table 1: Comparison of performance of CNN vs CNN with MLP on the test set.

Model	Group	AUC (95% CI)	Sensitivity	Specificity	NPV	PPV	F1 score
MR_MLP_CNN_AUG	>40	0.854, (0.768, 0.94)	0.762	0.777	0.102	0.99	0.18
	40-59	0.818, (0.76, 0.876)	0.746	0.819	0.155	0.986	0.257
	60-80	0.792, (0.765, 0.818)	0.779	0.68	0.148	0.977	0.249
	<80	0.734, (0.698, 0.769)	0.662	0.681	0.217	0.938	0.327
	males	0.823, (0.8, 0.847)	0.777	0.728	0.155	0.981	0.258
	females	0.776, (0.748, 0.804)	0.724	0.713	0.186	0.966	0.296
	Overall	0.805, (0.787, 0.822)	0.734	0.741	0.175	0.974	0.283

Table 2: Comparison of model performance over different population groups.

Conclusion

AI-based ECG models integrating CNN, MLP, and data augmentation techniques show potential for detecting moderate to severe MR. However, variability in performance across sex and age groups underscores the need for further optimization and validation to improve generalizability and reliability. While promising, these models should be viewed as supplementary diagnostic tools requiring additional refinement before broader clinical application.