

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

Tricuspid regurgitation (TR) is an underdiagnosed valvular heart disease associated with significant morbidity and mortality in advanced stages. Early detection of moderate to severe TR is critical to improving outcomes. With their accessibility and cost-effectiveness, electrocardiograms (ECGs) provide a promising platform for leveraging artificial intelligence (AI) to enhance TR detection.

Methods

Between 2009 and 2024, 51,788 adults [mean age 66 ± 16.5 years; 21,387 women (41%)] who underwent both echocardiography and ECG within a 14-day interval at Sheba Medical Center were identified. Moderate to severe TR was identified in 3,844 Echo-ECG pairs (7.4%). A Convolutional Neural Network (CNN) was trained on 31,072 samples (60%), validated on 10,359 samples (20%), and tested on 10,359 samples (20%), with stratification by TR severity, age, and sex. A Multi-Layer Perceptron (MLP) module was added to the CNN model, incorporating auxiliary data*, to enhance performance.

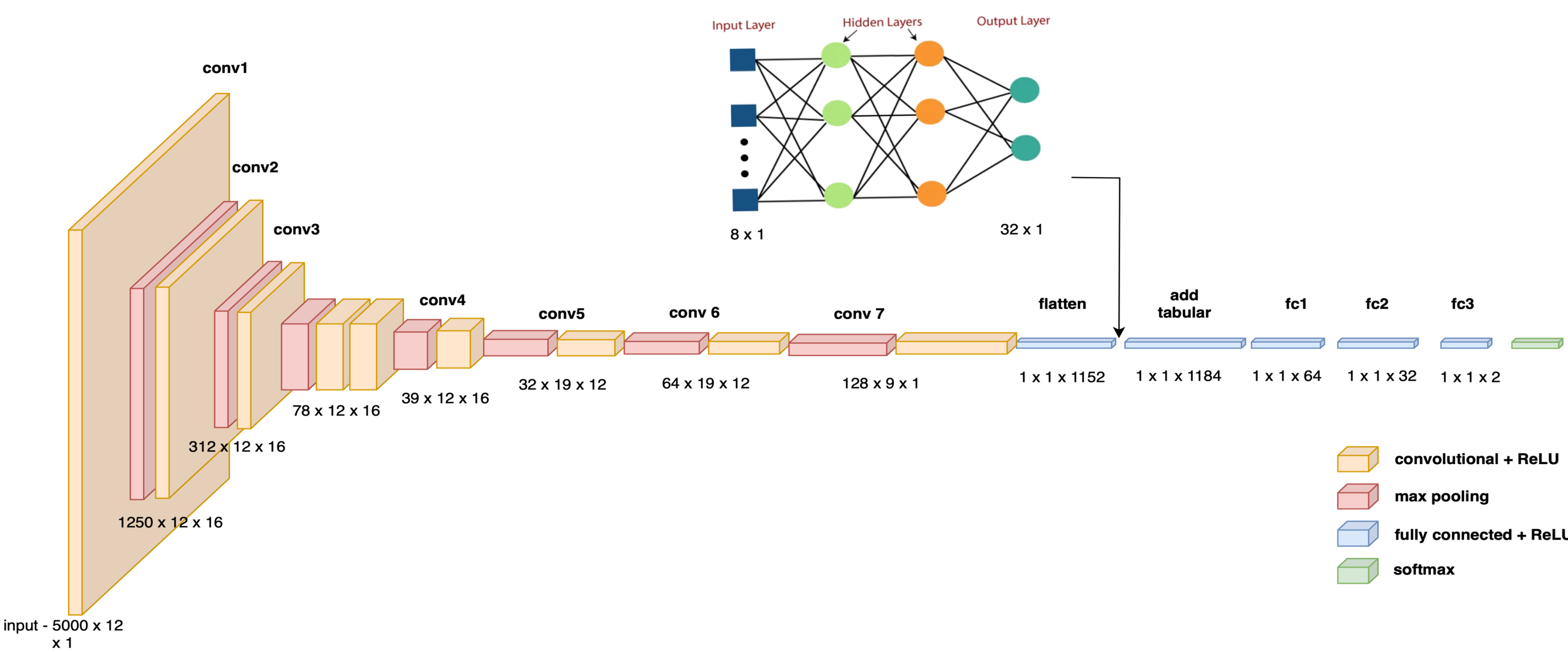


Image 1: Model Architecture

Results

The CNN achieved an Area Under the Curve (AUC) of 0.83, with sensitivity, specificity, and accuracy of 76.5%, 76.4%, and 76%, respectively. Adding the MLP module improved the AUC to 0.85, with corresponding metrics of 77.8%, 78.1%, and 77%. Subgroup analysis revealed the highest performance in patients aged 40–59 years (AUC 0.884), while lower AUCs were observed in patients under 40 years (AUC 0.758). Males demonstrated slightly better AUC (0.849) compared to females (0.832), with notable variability in sensitivity and specificity across subgroups. These results highlight both the potential and limitations of the model in accurately detecting TR in diverse populations.

Model	AUC	Accuracy	Sensitivity	Specificity
CNN	0.83	76%	76.5%	76.4%
CNN+MLP	0.85	77%	77.8%	78.1%

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

Model	Group	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	F1 score
TR_MLP_CNN	>40	0.866, (0.717, 1.0)	0.833	0.861	0.088	0.997	0.159
	40-59	0.884, (0.839, 0.929)	0.83	0.81	0.106	0.994	0.188
	60-80	0.836, (0.809, 0.862)	0.778	0.777	0.165	0.984	0.272
	<80	0.758, (0.724, 0.792)	0.691	0.727	0.273	0.941	0.391
	males	0.849, (0.824, 0.875)	0.792	0.763	0.128	0.988	0.22
	females	0.832, (0.809, 0.856)	0.817	0.723	0.214	0.977	0.339
	Overall	0.848, (0.832, 0.865)	0.778	0.781	0.184	0.982	0.298

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

Conclusion

AI-driven ECG models combining CNN and MLP demonstrate potential for detecting moderate to severe TR with reasonable accuracy. However, variability in performance across age and sex groups underscores the need for further refinement to improve generalizability and clinical utility. These findings suggest that AI-based tools could serve as supplementary aids in TR diagnosis, with further validation required before clinical integration.