# Deep Learning for ECG Screening of Moderate to Severe Aortic Valve Stenosis

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

## Aim and Background

Aortic stenosis (AS) is a significant health challenge where early diagnosis is crucial to improve outcomes. Advances in AS management emphasize timely detection to mitigate the risk of progression. Electrocardiograms (ECGs), as accessible and cost-effective diagnostic tools, have potential for AI-driven innovation in AS detection.

## Methods

From 2009 to 2024, we identified 53,220 adults [mean age 65.5 ± 16.6; 21,408 women (40.2%)] with echocardiography and ECG performed within 14 days at Sheba Medical Center. Moderate to severe AS was identified in 2,891 Echo-ECG pairs (5.4%). A Convolutional Neural Network (CNN) was trained on 37,253 samples (70%), validated on 5,323 (10%), and tested on 10,644 (20%), with stratification by AS severity, age, and sex. A Multi-Layer Perceptron (MLP) module was added to enhance performance.

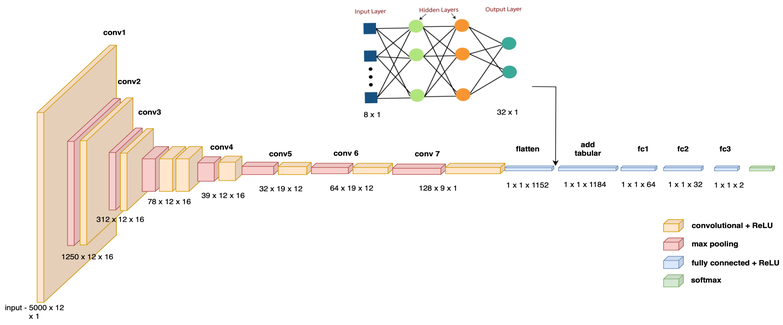


**Image 1: Model Architecture**

## Results

The CNN achieved an AUC of 0.78, with sensitivity, specificity, and accuracy of 71.2%, 71.04%, and 71%, respectively. Adding MLP improved the AUC to 0.84, with sensitivity, specificity, and accuracy increasing to 75.9%, 78.7%, and 78%. Excluding mild AS cases further enhanced performance, yielding an AUC of 0.85, sensitivity of 77.7%, specificity of 80.6%, and accuracy of 77%. Subgroup analysis showed consistent performance across sexes, with slightly higher metrics for females, and improved results in older age groups.

| Model | AUC | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|
| CNN | 0.78 | 71% | 71.2% | 71.04% |
| CNN + MLP | 0.84 | 78% | 75.9% | 78.7% |
| CNN + MLP + test exclusion | 0.85 | 77% | 77.7% | 80.6% |

**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 |
|---|---|---|---|---|---|---|---|
| CNN + MLP + test exclusion | >40 | 0.832, (0.68, 0.984) | 1 | 0.755 | 0.009 | 1 | 0.018 |
| | 40-59 | 0.733, (0.558, 0.907) | 0.6 | 0.79 | 0.015 | 0.997 | 0.029 |
| | 60-80 | 0.822, (0.78, 0.864) | 0.701 | 0.828 | 0.095 | 0.991 | 0.167 |
| | <80 | 0.789, (0.749, 0.829) | 0.721 | 0.761 | 0.237 | 0.964 | 0.357 |
| | males | 0.858, (0.823, 0.893) | 0.816 | 0.792 | 0.088 | 0.994 | 0.159 |
| | females | 0.863, (0.835, 0.892) | 0.741 | 0.834 | 0.15 | 0.988 | 0.249 |
| | Overall | 0.858, (0.835, 0.881) | 0.777 | 0.806 | 0.109 | 0.992 | 0.191 |

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

## Conclusion

AI-based ECG models incorporating CNN and MLP offer potential as supplementary tools for detecting moderate to severe AS, particularly when mild cases are excluded. While promising, the models require further refinement to achieve optimal accuracy and reliability across diverse patient populations.