Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

MobileViTv2 with a normalised cosine-similarity classification head and synthetic data balancing for ECG arrhythmia classification on the MIT-BIH Arrhythmia Database

César Nogueira Silva

cesar_n_silv@hotmail.com

Centro Universitário FEI

Leila Cristina Carneiro Bergamasco

Centro Universitário FEI

Research Article

Keywords: Electrocardiogram, Arrhythmia classification, MobileViTv2, Synthetic data, Continuous wavelet transform, Explainable artificial intelligence

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11472796/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose

To classify heartbeats from the MIT-BIH Arrhythmia Database into the five AAMI classes with a lightweight hybrid convolution–transformer network, while mitigating class imbalance and examining the model's errors.

Methods

Beats were segmented around the R peak, converted to continuous wavelet transform (CWT) images and used to train MobileViTv2 with its linear classifier replaced by a normalised cosine-similarity head (NormHead). The training set was balanced to 4,000 beats per class using beats synthesised by a Wasserstein GAN with gradient penalty (WGAN-GP) and a convolutional autoencoder; the test set (7,712 beats) contained only real beats. Grad-CAM was used for qualitative error analysis.

Results

On a stratified, beat-level hold-out set the model reached an accuracy of 0.9793 and a macro F1-score of 0.9538 (0.9795 and 0.9515 in a second run), against 0.9745 and 0.9418 for the same backbone with a standard head. Classes F and SVE remained the hardest (F1-scores of 0.8812 and 0.9267). Synthetic beats were highly similar to real beats in time and frequency for most classes.

Conclusion

The cosine-similarity head gave a modest improvement over a standard head in this setting. Because training and test beats were split at the beat level and not by patient, the results describe intra-patient performance and are not evidence of generalisation to unseen patients; patient-independent and cross-database validation are still required.

Introduction

Cardiovascular diseases are the leading cause of death worldwide, accounting for roughly 31% of global deaths (World Health Organization 2021). Arrhythmias, which are disturbances of the heart's rhythm or electrical conduction, can precede sudden cardiac death, so their early detection matters. The electrocardiogram (ECG) is the main tool for monitoring and diagnosing them, and wearable devices are making continuous ECG recording increasingly common.

Deep neural networks are widely used to classify ECG beats, including approaches that convert beats to continuous wavelet transform (CWT) images and process them with convolutional and recurrent

networks (Toma and Choi 2022), attention-based networks with explainability (Ardiyansyah and Mandala 2024), and patient-specific one-dimensional convolutional networks (Kiranyaz et al. 2016). Three issues still limit clinical use: the strong class imbalance of public databases, which biases models towards the majority class; the opacity of the models; and evaluation protocols that differ between studies. In particular, results obtained when beats from the same patient appear in both training and test data are not comparable with patient-independent (inter-patient) results, such as those of the protocol of de Chazal et al. (2004).

This study addresses imbalance and efficiency on the MIT-BIH Arrhythmia Database (Moody and Mark 2001; Goldberger et al. 2000) with the following elements: (i) balancing of the training set with synthetic beats generated by a Wasserstein GAN with gradient penalty (Gulrajani et al. 2017) and by a convolutional autoencoder; (ii) classification of CWT images with MobileViTv2 (Mehta and Rastegari 2022), a hybrid convolution–transformer network designed for mobile devices; (iii) replacement of the standard linear classifier by a normalised cosine-similarity head, here called NormHead, which follows the normalised-softmax and cosine-classifier idea used in face recognition (Wang et al. 2017, 2018) and applies it to ECG beats; and (iv) Grad-CAM (Selvaraju et al. 2020) for qualitative analysis of misclassified beats. We compare the proposed model with the same backbone using a standard head and with other architectures evaluated on the same test set.

The evaluation uses a stratified beat-level hold-out split, not a patient-independent split. The results therefore measure performance when beats of the same patients are seen in training, and the Discussion sets out the consequences of this choice. Earlier work by the first author evaluated ResNet50 on raw-beat and CWT images for four AAMI classes (Silva and Castro 2025) and classical classifiers on the same four classes (Silva et al. 2023). The present study differs in using five classes, synthetic balancing, MobileViTv2 and NormHead.

Methods

Data and pre-processing

The MIT-BIH Arrhythmia Database (Moody and Mark 2001; Goldberger et al. 2000) contains 48 half-hour two-channel ambulatory ECG recordings sampled at 360 Hz, with beat annotations made by specialists. The recordings were read with the Python wfdb package. Each beat was segmented around its annotated R peak as a window of 300 samples (149 before and 150 after the peak). The annotated beat types were grouped into five classes following ANSI/AAMI EC57 (ANSI/AAMI 2020), as shown in Table 1. Class D gathers the unclassifiable and paced beat types (f, /, Q) and corresponds to the AAMI class usually labelled Q.

Table 1
Mapping of MIT-BIH beat annotations to the five classes used in this study and number of real beats of each class in the test set

Class	MIT-BIH annotations	Test beats (real)
N (normal)	N, L, R, e, j	4,814
SVE (supraventricular ectopic)	A, a, J, S	557
VE (ventricular ectopic)	V, E	1,401
F (fusion)	F	161
D (unclassifiable / paced)	f, /, Q	779
Total		7,712

Each 300-sample beat was converted to a 224 × 224 image by the CWT with the Gaussian-8 mother wavelet (gaus8 in PyWavelets), which represents the beat in the time–frequency plane.

Evaluation protocol

Real beats were split into 80% for training and 20% for testing by stratified sampling on class and sub-class (the rare sub-class S was forced to appear in the test set). The split was made at the beat level, so beats of the same record may be in both partitions. The test set contains 7,712 real beats and no synthetic beats. Only the training partition was balanced. A separate validation partition was not used: the checkpoint with the highest accuracy on the held-out set was kept for the final evaluation, which makes the reported figures optimistic to some degree.

Synthetic beat generation and balancing

Synthetic beats were generated from raw, unfiltered beats so that noise and artefacts of the real signals were preserved. For classes with many beats, a WGAN-GP was trained (latent dimension 100, 300 epochs, Adam with learning rate 1×10^{-4} and $\beta = (0.5, 0.9)$, gradient-penalty weight 10, five critic updates per generator update). For classes with few beats (about 200 per class), a one-dimensional convolutional autoencoder was trained (latent dimension 16, 1,000 epochs, Adam with learning rate 1×10^{-3}, loss combining mean squared error, a spectral term with weight 0.3 and a smoothness term with weight 0.01), and new beats were obtained by adding Gaussian noise (standard deviation 0.05) to the latent vector. Both generators produced 300-sample beats.

The training set was balanced to exactly 4,000 beats per class. Classes N and VE were downsampled in proportion to their sub-classes; classes D, SVE and F were upsampled with synthetic beats, again preserving the proportion of sub-classes. Fidelity of the synthetic beats was assessed per class with four measures: the Pearson correlation between the mean real and mean synthetic beat (Corr_mean);

the correlation between normalised power spectra obtained by the fast Fourier transform (Corr_spectrum); the Wasserstein distance between those spectra; and the relative error of signal energy, computed only for classes with few real beats.

Proposed model

The backbone is MobileViTv2_100 from the timm library, initialised with ImageNet weights. Its linear classification layer was replaced by NormHead, which applies adaptive global average pooling to obtain the feature vector z, L2-normalises z and each class weight vector w_k, and takes their inner product as the logit of class k (Eqs. 1 and 2):

$\hat{z} = z / \|z\|_2$, $\hat{w}_k = w_k / \|w_k\|_2$ (1)

$\text{logit}_k = \hat{z} \cdot \hat{w}_k = \cos \theta_k$ (2)

The logit of each class is therefore the cosine of the angle between the beat representation and the class vector, and it does not depend on the norm of z. No scaling factor is applied to the logits. The network was trained for 30 epochs with AdamW (learning rate 3×10^{-4}, weight decay 10^{-4}), cosine-annealing learning-rate schedule, batch size 128, automatic mixed precision, and cross-entropy loss with label smoothing of 0.05 and class weights inversely proportional to class frequency in the training set. Random horizontal flipping was the only data augmentation.

The baseline was the same backbone with its standard linear head, initialised with ImageNet weights and trained with AdamW (learning rate 1×10^{-3}, weight decay 10^{-4}, cosine annealing, label smoothing 0.1, class weights, early stopping with patience of five epochs), followed by progressive fine-tuning in which only the last blocks and the classifier were unfrozen (learning rate 1×10^{-4}, 10 epochs). The training recipes of the baseline and of NormHead therefore differ in more than the head.

Comparison models

Other models were evaluated for comparison: ConvNeXt V2 (Woo et al. 2023); Swin Transformer V2 (Liu et al. 2022; swinv2_tiny_window8_256 from timm, 256 × 256 inputs, AdamW with learning rate 3×10^{-4}, 35 epochs); MobileViTv2 with focal loss ($\gamma = 2$) and decision thresholds; MobileNetV2 features reduced by principal component analysis and classified by an RBF support vector machine; a hierarchical scheme of a CNN and binary support vector machines; MobileNetV2 with a gated recurrent unit; and a MobileViT with a capsule network head. All used CWT images as input.

Metrics, timing and interpretability

Accuracy, per-class precision, recall (sensitivity) and F1-score, macro and weighted averages and the confusion matrix were computed with scikit-learn. For a screening view, predictions were also collapsed

into normal (class N) and arrhythmia (all other classes). Inference time per beat was measured on a GPU for the network only. It does not include beat segmentation or CWT computation. For the baseline, the total inference time over the test set was divided by the number of beats; for the NormHead model, the mean over 100 consecutive single-beat inferences with explicit GPU synchronisation was used. Grad-CAM was computed on convolutional layers of the backbone and applied to misclassified beats.

Signal extraction, beat segmentation and synthetic-beat generation were run on a local workstation (Dell Inspiron 14 5440, Intel processor at about 1.8 GHz, 16 GB of RAM, Windows 11, Python 3.13.9). The classification models were trained and evaluated on Google Colab with an NVIDIA A100 GPU.

Use of generative AI

A generative AI assistant was used to draft and edit the text of this manuscript from the first author's Master's dissertation. It was not used for the research design, the experiments, the data analysis or the interpretation of the results. The authors reviewed all text and take full responsibility for it.

Results

Fidelity of synthetic beats

Table 2 shows the fidelity measures. The mean beat of most classes correlated with the real mean beat above 0.95, and spectral correlations were mostly above 0.99 with Wasserstein distances below 0.003. There were exceptions: classes S (Corr_mean 0.9008) and Q (0.9213) had lower temporal correlation, class e had a spectral correlation of only 0.4086 and a Wasserstein distance of 0.0085, and class J had a relative energy error of 0.2814. These classes have few real beats, which limits what the generators can learn.

Table 2
Similarity between real and synthetic beats by MIT-BIH class

Class	Corr_mean	Corr_spectrum	Wasserstein (spectrum)	Relative energy error
V	0.9617	0.9997	0.0002	–
L	0.9894	0.9912	0.0015	–
R	0.9866	0.9999	0.0003	–
A	0.9589	0.9998	0.0007	–
F	0.9950	0.9506	0.0011	–
j	0.9609	0.9956	0.0024	–
a	0.9543	0.9997	0.0004	0.0557
E	0.9966	0.9996	0.0001	0.0013
J	0.9922	0.9999	0.0000	0.2814
e	0.9900	0.4086	0.0085	0.0569
Q	0.9213	0.9624	0.0027	0.1377
S	0.9008	0.9994	0.0003	0.0101
f	0.9907	0.9994	0.0003	0.0298
/	0.9955	0.9881	0.0008	0.0252

Relative energy error was computed only for classes with few real beats (about 200 per class). Class N was not synthesised.

Performance of the proposed model

On the 7,712 real test beats, MobileViTv2 with NormHead reached an accuracy of 0.9793 and a macro F1-score of 0.9538 (Table 3, Fig. 1). Class D was almost perfectly separated (F1-score 0.9968), which raises the macro-averaged metrics. Class F had the lowest F1-score (0.8812). Most errors were between N and SVE: 41 N beats were predicted as SVE and 31 SVE beats as N. A second training run of the same model gave an accuracy of 0.9795 and a macro F1-score of 0.9515.

Table 3
Classification report of MobileViTv2 with NormHead on the held-out test set (first run)

Class	Precision	Recall	F1-score	Support
D	0.9962	0.9974	0.9968	779
F	0.8868	0.8758	0.8812	161
N	0.9865	0.9857	0.9861	4,814
SVE	0.9234	0.9300	0.9267	557
VE	0.9779	0.9786	0.9782	1,401
Macro average	0.9541	0.9535	0.9538	7,712
Weighted average	0.9793	0.9793	0.9793	7,712
Accuracy			0.9793	7,712

Test set: 7,712 real beats, 20% stratified hold-out at the beat level.

Collapsing predictions into normal versus arrhythmia, a second run of the CWT model (accuracy 0.9795) made 54 false positives and 67 false negatives, so 67 of the 2,898 arrhythmic beats (2.3%) were labelled normal.

Effect of the classification head

Table 4 compares the standard-head baseline with NormHead on the same backbone, data and test set. NormHead was higher in accuracy (+ 0.0048 in the first run) and macro F1-score (+ 0.0120), with the largest gains in classes F (F1-score + 0.0325) and SVE (+ 0.0204), whereas macro-averaged recall was slightly lower (0.9535 against 0.9549). The two NormHead runs differed by 0.0002 in accuracy and 0.0023 in macro F1-score, so the difference between the heads is larger than the observed run-to-run variation. This is however a comparison of single baseline and two NormHead runs with different training recipes (see Methods), and it is not an ablation of the other components (CWT input, synthetic balancing).

Table 4
Standard linear head and NormHead on MobileViTv2 (same test set of 7,712 real beats)

Metric	Standard head	NormHead, run 1	NormHead, run 2
Accuracy	0.9745	0.9793	0.9795
Macro precision	0.9298	0.9541	0.9546
Macro recall	0.9549	0.9535	0.9485
Macro F1-score	0.9418	0.9538	0.9515
F1-score, D	0.9974	0.9968	0.9981
F1-score, F	0.8487	0.8812	0.8734
F1-score, N	0.9838	0.9861	0.9874
F1-score, SVE	0.9063	0.9267	0.9214
F1-score, VE	0.9729	0.9782	0.9771

Comparison with other models

Table 5 lists models evaluated on the same 7,712-beat test set. NormHead had the highest accuracy and macro F1-score of the models listed, but the margins over the best alternatives are small (accuracy 0.9793 against 0.9773 for MobileNetV2 with PCA and an SVM; macro F1-score 0.9538 against 0.9487) and each model was run once. ConvNeXt V2 had a precision of only 0.6054 for class SVE. Swin Transformer V2 reached an accuracy of 0.9668 and a macro F1-score of 0.9667, but on a different, class-balanced test partition of 4,000 beats (800 per class) that included synthetic beats. It is therefore not comparable with Table 5 and was left out of it; its macro F1-score on that partition is higher than that of the NormHead model on the real test set, and re-evaluating it on the real test set is needed before any ranking is drawn. Comparison with published results is not attempted, because published studies use different protocols, particularly with respect to patient separation.

Table 5
Models evaluated on the same held-out test set of 7,712 real beats, ordered by accuracy

Model	Accuracy	Macro F1-score
MobileViTv2 + NormHead (run 1)	0.9793	0.9538
MobileNetV2 + PCA + SVM (RBF)	0.9773	0.9487
MobileViTv2 + capsule network	0.9759	0.9439
MobileViTv2 + focal loss + thresholds	0.9758	0.9446
MobileViTv2, standard head (baseline)	0.9745	0.9418
MobileNetV2 + GRU	0.9669	0.9299
Hierarchical CNN + SVMs	0.9422	0.8899
ConvNeXt V2	0.9375	0.8873

Input representation

With the same NormHead model trained on the raw beat drawn as a time-domain image, without CWT, accuracy was 0.9751 and macro F1-score 0.9469, against 0.9793 and 0.9538 with CWT images. In the screening view this model made 78 false positives and 73 false negatives. The time-domain input still requires converting each beat to an image, so it does not remove the image-generation step.

Inference time

Table 6 gives the measured GPU inference times. The 1.67 ms per beat reported for the baseline MobileViTv2 is the total inference time over the test set (12.88 s) divided by the number of beats. The NormHead model was measured differently (single-beat inference, mean of 100 runs with GPU synchronisation) and gave 12.75 ms. The two values are not comparable, and no timing was made on an embedded or wearable processor or for the CWT computation.

Table 6
Mean inference time per beat on a GPU (network only; measurement protocols differ between rows)

Model	Time per beat (ms)
MobileViTv2 + capsule network	0.99
MobileViTv2 + focal loss	1.57
MobileViTv2, standard head (baseline)	1.67
MobileNetV2 + GRU	2.00
MobileNetV2 + PCA + SVM	2.08
CNN + binary SVM	5.20
MobileNetV2	7.43
MobileViTv2 + NormHead	12.75
ConvNeXt V2	21.20
Swin Transformer V2	28.50

GPU: NVIDIA A100 (Google Colab). Beat segmentation and CWT computation are not included.

Qualitative analysis with Grad-CAM

Figure 2 shows Grad-CAM maps for five misclassified beats (three N beats predicted as SVE, one SVE predicted as N and one F predicted as N). Activation is localised, coarse and differs between examples. No quantitative evaluation of the maps was made, and the maps were not compared with annotated waves or noise, so they do not show which waveform components drive the predictions. We therefore use them only to illustrate individual errors.

Discussion

MobileViTv2 with NormHead classified the five AAMI classes of beats with an accuracy near 0.98 and a macro F1-score near 0.95 when beats of the same patients could occur in both training and test data. Replacing the linear head with a cosine-similarity head improved accuracy and macro F1-score modestly in our experiments, mainly through the minority classes F and SVE. Because a normalised head removes the dependence on feature norm, it can plausibly reduce the effect of amplitude and energy variations, but we did not test this mechanism, and the idea itself is not new in machine learning (Wang et al. 2017, 2018); the contribution here is its evaluation for ECG beat classification. Errors were concentrated between N and SVE, which is expected because supraventricular beats keep a narrow QRS complex and differ from normal beats mainly in the P wave and timing, which carry little energy in the CWT image.

Several limitations restrict the conclusions. First and most important, training and test beats were separated at random, not by patient, and the patient-independent protocol of de Chazal et al. (2004) was not applied. Cardiac morphology is patient-specific, so performance on unseen patients is likely to be lower than reported here, and synthetic beats derived from a patient's beats may resemble held-out beats of the same patient. Second, no separate validation set was used and the checkpoint was chosen on the test set, which biases the figures upwards. Third, each configuration was run once or twice, with different training recipes for the baseline and NormHead, so small differences between models should not be over-interpreted. Fourth, only one database was used, and the real beats used here are a subset of the database. Fifth, class D is almost perfectly separated and raises macro-averaged metrics. Sixth, inference times were measured on a data-centre GPU with different protocols, exclude CWT computation and say nothing about embedded or wearable devices, for which the cost of the CWT may dominate. Seventh, Grad-CAM was used only qualitatively. Finally, the synthetic beats of classes with few real beats are less faithful, and no experiment isolates the contribution of the synthetic data.

Future work should (i) evaluate the models with patient-independent splits, with validation and test data kept separate; (ii) repeat the experiments with several random seeds and run a full ablation of the CWT input, synthetic balancing and head with identical training recipes; (iii) test generalisation on independent databases; (iv) compare with end-to-end one-dimensional networks operating directly on the signal; and (v) measure the full pipeline, including CWT, on embedded hardware.

Conclusion

A MobileViTv2 with a normalised cosine-similarity head, trained on CWT images with a training set balanced by synthetic beats, classified AAMI beat classes of the MIT-BIH Arrhythmia Database with an accuracy of 0.9793 and a macro F1-score of 0.9538 on a beat-level hold-out set, a modest improvement over the same backbone with a standard head. These results describe intra-patient performance. Patient-independent evaluation, cross-database validation and measurement of the full pipeline on embedded hardware are needed before conclusions about clinical or wearable use can be drawn.

Declarations

Acknowledgments

The authors thank Centro Universitário FEI for institutional support.

Funding No funding was received to conduct this study. The first author received a merit scholarship from Centro Universitário FEI for his Master's studies, which was not linked to this research project.

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Ethics approval This study used only the publicly available, de-identified MIT-BIH Arrhythmia Database; no new data were collected from human participants.

Consent to participate and to publish Not applicable.

Data availability The MIT-BIH Arrhythmia Database is publicly available from PhysioNet (https://doi.org/10.13026/C2F305). The synthetic beats generated in this study are available from the corresponding author on reasonable request.

Code availability The code is available from the corresponding author on reasonable request.

Author contributions César Nogueira Silva: conceptualisation, methodology, software, investigation, formal analysis, writing – original draft. Leila Cristina Carneiro Bergamasco: supervision, writing – review and editing. Both authors read and approved the final manuscript.

Origin of the work This manuscript is based on the first author's Master's dissertation (Centro Universitário FEI, 2026), supervised by the second author.

References

1. ANSI/AAMI. ANSI/AAMI EC57:2012/(R)2020 Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms. Arlington: Association for the Advancement of Medical Instrumentation; 2020.
2. Ardiyansyah A, Mandala S. (2024) Accurate model for arrhythmia classification using hybrid attention-based convolutional neural network (ACNN) and gated recurrent units (GRU) with explainable artificial intelligence (XAI). In: 2024 12th International Conference on Information and Communication Technology (ICoICT), pp 57–64. https://doi.org/10.1109/ICoICT61617.2024.10698622
3. de Chazal P, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Trans Biomed Eng. 2004;51(7):1196–206. https://doi.org/10.1109/TBME.2004.827359.
4. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):e215–20. https://doi.org/10.1161/01.CIR.101.23.e215.
5. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A. (2017) Improved training of Wasserstein GANs. In: Advances in Neural Information Processing Systems 30. https://arxiv.org/abs/1704.00028
6. Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. IEEE Trans Biomed Eng. 2016;63(3):664–75. https://doi.org/10.1109/TBME.2015.2468589.
7. Liu Z, Hu H, Lin Y, Yao Z, Xie Z, Wei Y, Ning J, Cao Y, Zhang Z, Dong L, Wei F, Guo B. (2022) Swin Transformer V2: scaling up capacity and resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 11999–12009. https://doi.org/10.1109/CVPR52688.2022.01170

8. Mehta S, Rastegari M. (2022) Separable self-attention for mobile vision transformers. arXiv:2206.02680. https://doi.org/10.48550/arXiv.2206.02680
9. Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng Med Biol Mag. 2001;20(3):45–50. https://doi.org/10.1109/51.932724.
10. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis. 2020;128:336–59. https://doi.org/10.1007/s11263-019-01228-7.
11. Silva C, Lopes F, Matos J, Castro M. (2023) Arrythmia classification using MATLAB Classification Learner App. In: Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2023), vol 4: BIOSIGNALS, pp 220–225. https://doi.org/10.5220/0011666300003414
12. Silva CN, Castro MCF. (2025) Arrythmia classification using deep learning. In: XXIX Brazilian Congress on Biomedical Engineering, vol 3: Biomedical Informatics, and Biomedical Signal and Image Processing. IFMBE Proceedings. Springer Nature Switzerland, pp 462–473. https://doi.org/10.1007/978-3-031-94934-0_48
13. Toma TI, Choi S. A parallel cross convolutional recurrent neural network for automatic imbalanced ECG arrhythmia detection with continuous wavelet transform. Sensors. 2022;22(19):7396. https://doi.org/10.3390/s22197396.
14. Wang F, Xiang X, Cheng J, Yuille AL. (2017) NormFace: L2 hypersphere embedding for face verification. In: Proceedings of the 25th ACM International Conference on Multimedia, pp 1041–1049. https://doi.org/10.1145/3123266.3123359
15. Wang H, Wang Y, Zhou Z, Ji X, Gong D, Zhou J, Li Z, Liu W. (2018) CosFace: large margin cosine loss for deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 5265–5274. https://doi.org/10.1109/CVPR.2018.00552
16. Woo S, Debnath S, Hu R, Chen X, Liu Z, Kweon IS, Xie S. (2023) ConvNeXt V2: co-designing and scaling ConvNets with masked autoencoders. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 16133–16142. https://doi.org/10.1109/CVPR52729.2023.01548
17. World Health Organization. (2021) Cardiovascular diseases (CVDs). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Accessed 22 Oct 2026.

Figures

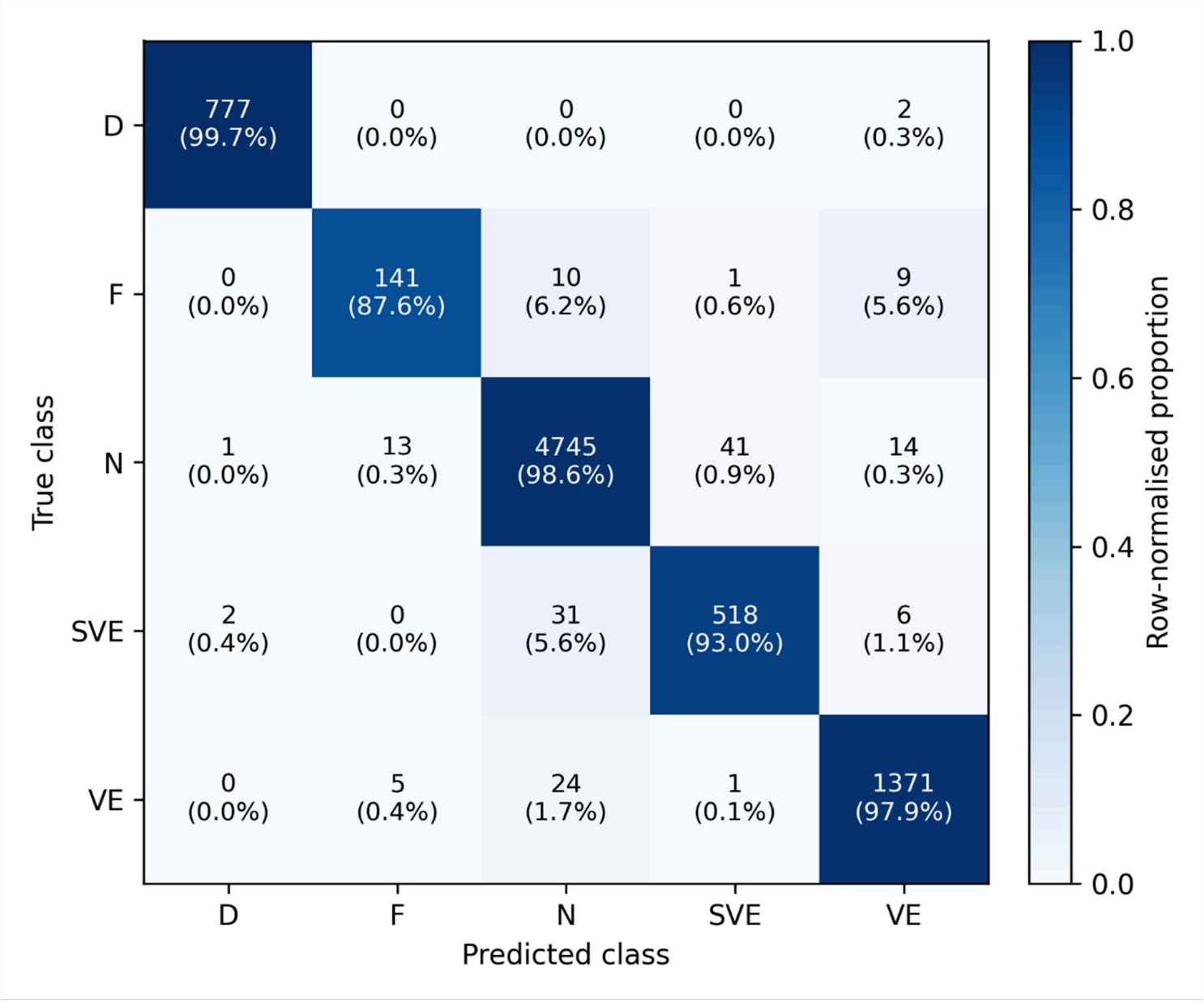


Figure 1

Confusion matrix of MobileViTv2 with NormHead on the held-out test set of 7,712 real beats (counts with row-normalised percentages in parentheses)