

GIRAFFE: A GENETIC PROGRAMMING ALGORITHM TO BUILD DEEP LEARNING ENSEMBLES FOR ARRHYTHMIA CLASSIFICATION

Damian Kucharski¹, Agata M. Wijata¹, Lu Fu², Weidong Lin², Yumei Xue²,
Jacek Kawa¹, Yalin Zheng^{3,4}, Gregory Y. H. Lip^{3,4} and Jakub Nalepa¹

¹Silesian University of Technology, Gliwice, Poland

²Department of Cardiology, Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital Guangdong Academy of Medical Sciences, Guangzhou, China

³Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool John Moores University and Liverpool Heart & Chest Hospital, Liverpool, United Kingdom

⁴Department of Eye and Vision Science, Institute of Life Course and Medical Sciences, University of Liverpool, Liverpool, United Kingdom

⁵Danish Center for Health Services Research, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark.

Contact: damian.kucharski@polsl.pl, jakub.nalepa@polsl.pl



Full paper



Background and Objectives

Cardiovascular diseases (CVDs) remain one of the leading causes of death worldwide. Cardiac arrhythmia, characterized by an irregular heartbeat, is one of the most frequently diagnosed CVDs. Electrocardiography is a standard tool to examine the heartbeat and diagnose arrhythmia. In this work we aim to produce genetic-programming based classification ensembles of deep learning models for accurate and robust arrhythmia classification from paper ECG scans.

Materials and Methods

Dataset

In this study, we use the **Guangzhou Heart Study** dataset [1] containing **1172 paper 12-lead ECG scans**. The dataset is split into non-overlapping training and test subsets (**1000 and 172 cases**) with **80% and 78% of patients presenting healthy signals**, respectively. Details are visible in **Table 1**.

Subset→	Training (T)	Test (Ψ)
Number of patients (% abnormal)	1000 (20%)	172 (22%)
Sex		
Female	645 (64.5%)	128 (74.4%)
Male	355 (35.5%)	44 (25.6%)
Age (years)		
Minimum	49	60
Mean	71	72
Standard deviation	6	6
Median	70	70
Maximum	96	91

Table 1: Characteristics of the Training (T) and test (Ψ) subsets of the Guangzhou Heart Study dataset utilized in this study

Genetic programming algorithm

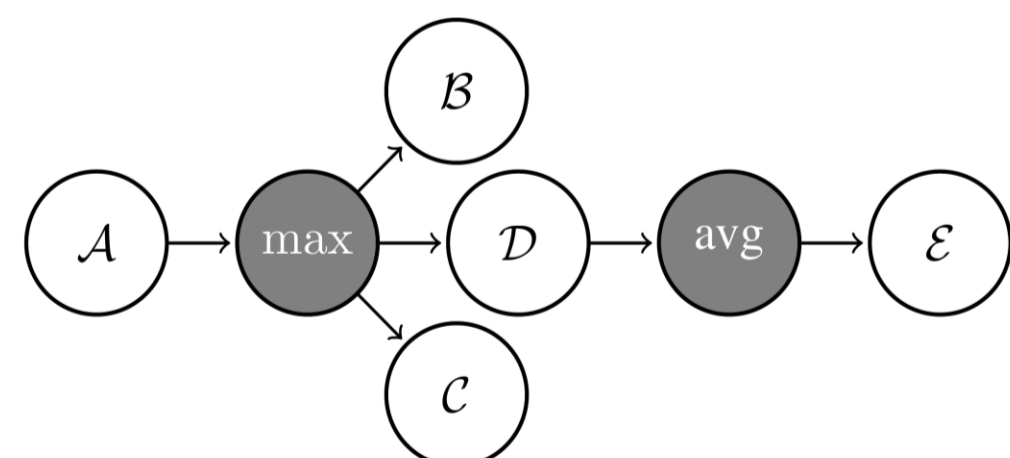
GIRAFFE benefits from mechanisms of genetic programming (GP) to evolve architecture, content and fusion scheme of the ensemble classifiers. Each tree (ensemble) is composed of two types of nodes.

- **Operation nodes (ON)**, representing any reduction operation, including the calculation of mean, minimum and maximum of inputted numerical values.
- **Value nodes (VN)**, storing the base models (one model in each VN), alongside the tree evaluations at the node level

Evolution process

The evolution process is shown in Figure 2. The model population, at first composed of single node base learner trees is evolved over g generations. New individuals are created by means of crossover (mixing of two tree structures) and pointwise mutations. Figure 1 shows example of computational tree and presents the evaluation process.

(a) An evolved ensemble classifier



(b) The node values and the results of their evaluation

Node ID	Value	Evaluation
A	{1.0, 2.0}	{3.0, 3.5}
B	{3.0, 2.5}	{3.0, 2.5}
C	{-1.0, 3.0}	{-1.0, 3.0}
D	{5.0, 4.0}	{3.0, 3.5}
E	{1.0, 3.0}	{1.0, 3.0}

Figure 1: Example of computational tree (an evolved ensemble), together with values of all nodes and the outcomes of evaluation process.

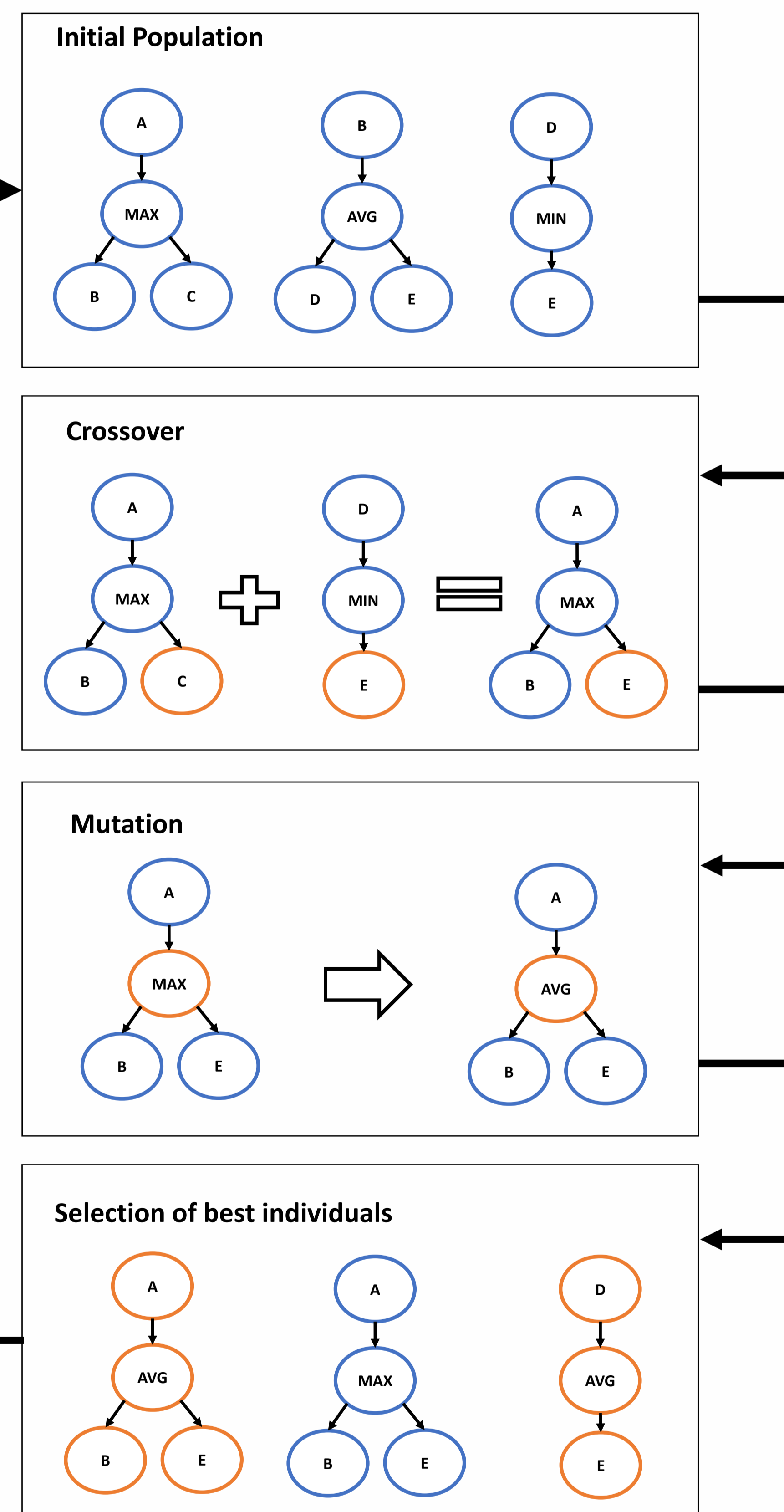


Figure 2: Illustration of the GIRAFFE's evolution process

Method	Rank	Validation set (V)							Test set (Ψ)						
		ROC-AUC	PR-AUC	Pt	Re	F1	Spe	MCC	ROC-AUC	PR-AUC	Pt	Re	F1	Spe	MCC
GIRAFFE	Top-1	0.994	0.969	0.958	0.920	0.939	0.993	0.929	0.960	0.940	0.897	0.921	0.909	0.970	0.883
	Top-2	0.994	0.968	0.958	0.920	0.939	0.993	0.929	0.948	0.913	0.886	0.816	0.849	0.970	0.810
	Top-3	0.994	0.967	0.958	0.920	0.939	0.993	0.929	0.945	0.920	0.889	0.842	0.865	0.970	0.828
	Top-4	0.994	0.966	0.958	0.920	0.939	0.993	0.929	0.951	0.929	0.868	0.868	0.868	0.963	0.831
	Top-5	0.993	0.961	0.958	0.920	0.939	0.993	0.929	0.954	0.917	0.842	0.842	0.842	0.955	0.797
Full ensemble	Top-1	0.997	0.985	0.920	0.920	0.920	0.986	0.906	0.965	0.944	0.850	0.895	0.872	0.955	0.835
	Top-2	0.943	0.767	0.821	0.920	0.868	0.966	0.846	0.914	0.731	0.791	0.895	0.840	0.933	0.793
	Top-3	0.943	0.767	0.821	0.920	0.868	0.966	0.846	0.914	0.731	0.791	0.895	0.840	0.933	0.793
	Top-4	0.943	0.767	0.821	0.920	0.868	0.966	0.846	0.901	0.711	0.786	0.868	0.825	0.933	0.774
	Top-5	0.997	0.985	0.774	0.960	0.857	0.952	0.836	0.965	0.944	0.761	0.921	0.833	0.918	0.786
Random	Top-1	0.981	0.931	0.880	0.880	0.880	0.979	0.859	0.959	0.941	0.805	0.868	0.835	0.940	0.787
	Top-2	0.976	0.931	0.880	0.880	0.880	0.979	0.859	0.958	0.943	0.833	0.921	0.875	0.948	0.839
	Top-3	0.977	0.925	0.880	0.880	0.880	0.979	0.859	0.954	0.927	0.805	0.868	0.835	0.940	0.787
	Top-4	0.976	0.920	0.880	0.880	0.880	0.979	0.859	0.955	0.934	0.800	0.842	0.821	0.940	0.768
	Top-5	0.973	0.918	0.880	0.880	0.880	0.979	0.859	0.954	0.929	0.786	0.868	0.825	0.933	0.774
Base model	Top-1	0.992	0.954	0.913	0.840	0.875	0.986	0.856	0.938	0.884	0.844	0.711	0.771	0.963	0.718
	Top-2	0.957	0.901	0.913	0.840	0.875	0.986	0.856	0.982	0.943	0.900	0.711	0.794	0.978	0.752
	Top-3	0.977	0.929	0.952	0.800	0.870	0.993	0.854	0.967	0.933	0.909	0.789	0.845	0.978	0.808
	Top-4	0.983	0.942	0.846	0.880	0.863	0.972	0.839	0.961	0.921	0.938	0.789	0.857	0.985	0.826
	Top-5	0.977	0.929	0.846	0.880	0.863	0.972	0.839	0.967	0.933	0.882	0.789	0.833	0.970	0.791

Table 2: Comparison of GIRAFFE generated ensembles to alternative ensembling methods, random selection and base models. Set-wide best metrics are **boldfaced**, whereas the second best are underlined.

Base learners

We have utilized several deep learning architectures to promote diversity in base learners. The architectures include *ResNet*[2], *GoogLeNet*[3] and *InceptionV3*[4].

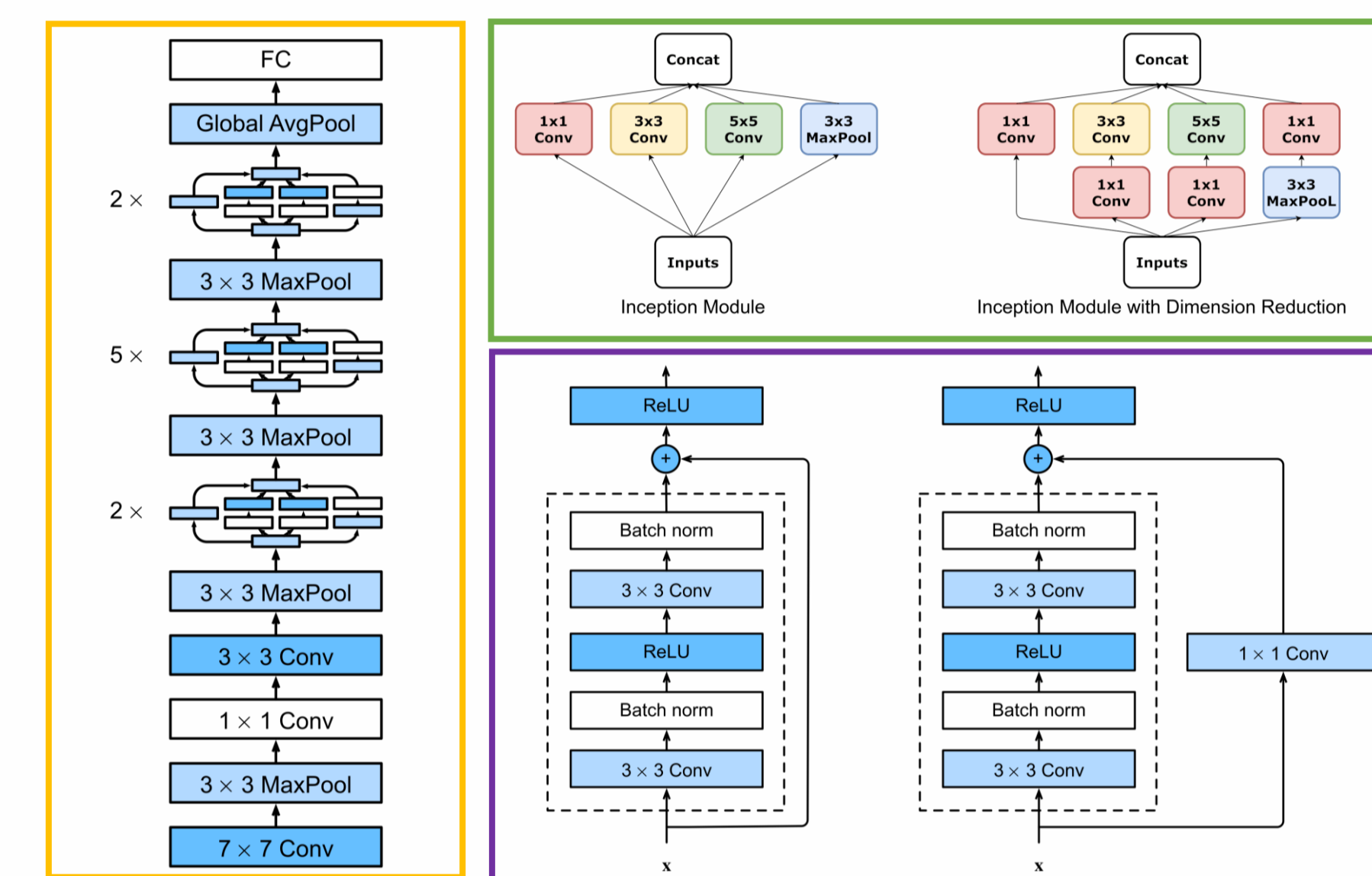


Figure 3: Architectures utilized as base learners. Figure shows GoogLeNet[3] (yellow), Inception modules[4] (green) and ResNet Blocks[2] (purple).

Results and Conclusions

Based on experimental results we conclude that:

- **GIRAFFE performs model selection** and finds relevant base learners. Five out of 56 models were selected (9%). (Figure 3)
- **GIRAFFE emphasizes the most important models**. Some models may be included multiple times in the calculation, emphasizing their importance. M_2 was selected twice. (Figure 5)
- **GIRAFFE outperforms alternative ensembling techniques**. Compared to other techniques like voting, averaging, stacking and random selection GIRAFFE obtained the most balanced and highest results across multiple classification metrics. (Table 2)
- **Models closer to the root typically have the greatest influence on the final prediction**, which can lead to a scenario where a single base learner makes the final decision. (Figure 4)

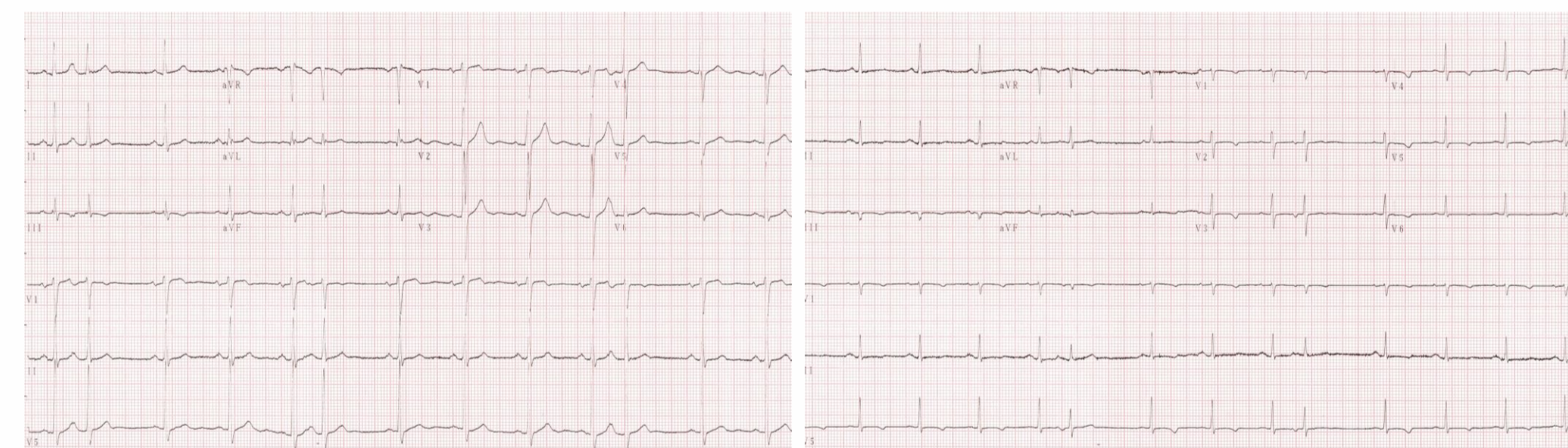
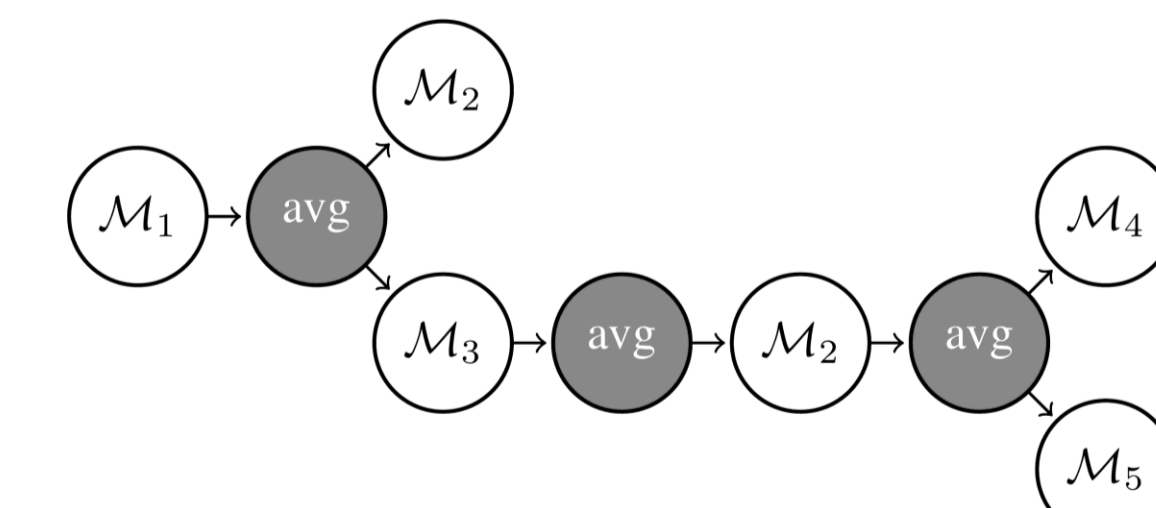


Figure 4: Examples of Ψ images, where the ensemble evolved by GIRAFFE provided: the correct prediction (all other Top-1 models delivered incorrect prediction) - left, and the incorrect prediction (predicting patient of arrhythmia instead of the healthy patient, with four its base models (apart from M1) delivering the correct one - right).

(a) The best ensemble classifier evolved using GIRAFFE



(b) The details of the base models (M_1 - M_5)

Model	Architecture	Only head training	Fold	MCC (over Ψ)
M_1	GoogLeNet	no	4	0.771
M_2	IADM	no	5	0.718
M_3	GoogLeNet	no	3	0.738
M_4	Resnet18	yes	3	0.452
M_5	IADM	no	T	0.826

Figure 5: (a) The best ensemble classifier evolved using GIRAFFE, together with the (b) details of its base classifiers and exploited operations. The VNs and ONs are rendered in gray and white.

References

[1] H. Deng et al., "Epidemiological characteristics of atrial fibrillation in Southern China: Results from the Guangzhou Heart Study," *Scientific Reports*, vol. 8, no. 1, pp. 17829, Dec. 2018, Number: 1 Publisher: Nature Publishing Group.

[2] K. He et al., "Deep residual learning for image recognition," Dec. 2015, arXiv:1512.03385 [cs].

[3] C. Szegedy et al., "Going deeper with convolutions," Sept. 2014, arXiv:1409.4842 [cs].

[4] C. Szegedy, "Rethinking the inception architecture for computer vision," Dec. 2015, arXiv:1512.00567 [cs].