Deep Learning Optimisation for Sports Cardiology: Neural Architecture Search-Driven Arrhythmia Classification

Erik Vanegas Müller[1,2], Arese Joe-Oshodi[2], Liang He[1,2], Abhirup Banerjee[2,3], Mauricio Villarroel[1,2]

[1]The Podium Institute for Sports Medicine and Technology, University of Oxford, United Kingdom
[2]Institute of Biomedical Engineering, University of Oxford, United Kingdom
[3]Division of Cardiovascular Medicine, University of Oxford, United Kingdom

Abstract

Athletes develop cardiac adaptations from training, but distinguishing between physiological and pathological changes remains a challenging clinical problem. We leverage Neural Architecture Search (NAS) to explore deep learning architecture optimisation for cardiac rhythm classification in professional football athletes. We developed a baseline and three pathophysiology-aware (rate, depolarisation, repolarisation) neural networks on a dataset from the general population and tested them on a sports dataset. Rate-based arrhythmias needed larger temporal windows (kernel size = 21) for extended pattern capture, whilst depolarisation abnormalities required small kernels (kernel size = 3) for high-frequency morphological features with maximum attention heads (number of heads = 16) for parallel feature analysis. Repolarisation disorders showed intermediate complexity (kernel size = 9). All categories required high regularisation (dropout rates between 0.5–0.7). While the baseline model achieved the highest overall averaged area under the receiver operating characteristic curve of 0.62, individual rhythm-specific performance suggests that specialisation benefits are present. Pathophysiology-informed NAS designs represent a step toward domain-informed arrhythmia detection in sports cardiology.

1. Introduction

Most arrhythmias during sports activities are due to undetected structural cardiac defects or channelopathies in which high-intensity exercise can trigger abnormal cardiac rhythms [1]. These cardiac defects resemble physiological and anatomical adaptations associated with sports [2]. Sinus Bradycardia (SB), Incomplete Right Bundle Branch Block (IRBBB), and T-wave Inversion (TWI) are common cardiac adaptations in football athletes [3]. Up to one in every 300 athletes may have an arrhythmic substrate that predisposes them to sudden cardiac arrest [4]. Distinguishing between physiological and pathological cardiac adaptations in athletes is a challenging clinical problem.

Neural Architecture Search (NAS) is an emerging Artificial Intelligence (AI) research direction that automatically designs optimal neural network (NN) architectures for specific tasks, including arrhythmia classification. N. Fayyazifar [5] used NAS to optimise a convolutional NN structure for the detection of Atrial Fibrillation (AF) from single-lead electrocardiogram (ECG) data using the Efficient Neural Architecture Search algorithm. Liu et al. [6] explored optimal attention architecture through NAS to improve pattern learning of temporal and lead-associated information. Asadi et al. [7] utilised NAS to automatically design a convolutional NN for the classification of paroxysmal AF using the Differentiable ArchiTecture Search (DARTS).

We leverage NAS to explore pathophysiology-aware NN architecture optimisation in professional football athletes. We design a custom-made search space that focuses on the initial kernel size (physical interpretability), dropout rate (regularisation), and the number of attention heads (simultaneous feature analysis). We split a heterogeneous dataset into pathophysiology-specific datasets to develop our proposed NAS. Our goal is to determine the extent to which pathophysiology-aware training improves the classification of sports-related cardiac arrhythmias.

2. Methods

2.1. Datasets

We use two datasets for developing the NAS-driven arrhythmia classification: the PhysioNet Challenge 21 [8] dataset for training and validation, and the Pro-Football 12-lead Resting Electrocardiogram Database (PF12RED) [9] for testing.

We first split the PhysioNet Challenge 21 dataset

(n = 88,253) into a baseline dataset (all ECGs), and three pathophysiology datasets: one with rate-based cardiac rhythms (i.e., timing and regularity abnormalities), one with ventricular depolarisation abnormalities (i.e., QRS-related abnormalities), and one with ventricular repolarisation pathologies (i.e., ST-T related abnormalities).

The Rate NAS dataset (approximately $35.9\,\%$ of the PhysioNet Challenge 21 dataset) consists of Bradycardia, Sinus Arrhythmia, SB, and Sinus Tachycardia. The dataset for ventricular depolarisation ($27.1\,\%$) contains Bundle Branch Block (BBB), Complete Left BBB, Complete Right BBB, 1st Degree AV Block, IRBBB, Left Anterior Fascicular Block, Left Axis Deviation, Left BBB, Low QRS Voltage, Non-specific Intraventricular Conduction Block, Poor R Wave Progression, Prolonged PR Interval, Q Wave Abnormal, Right Axis Deviation, and Right BBB. Finally, the ventricular repolarisation dataset ($15.8\,\%$) comprises T-wave abnormalities and TWI. Nine arrhythmias do not fit into the three pathological awareness categories, including supraventricular (AF, Atrial Flutter, Premature Atrial Contraction, Supraventricular Premature Beats) and ventricular (Premature Ventricular Contractions, Ventricular Premature Beats) arrhythmias, pacing rhythms, mixed pathophysiology (Long QT), and Normal Sinus Rhythm (NSR).

The PF12RED is an open-source dataset from Spanish professional football players containing 161 resting ECGs with the cardiac rhythms of SB, IRBBB, and TWI, in addition to NSR. In relation to NSR, SB is a rate problem, while IRBBB and TWI are ventricular depolarisation and repolarisation abnormalities, respectively.

2.2. Neural Network Architecture

We propose a NAS-driven arrhythmia classification based on rate and ventricular depolarisation and repolarisation. Our NN is based on a residual NN with a multi-head attention mechanism, as described in [10]. The input goes through a convolution head with a kernel size of 1×15, while using a stride of 2. The embedding dimension has an output of 256 channels and undergo a batch normalisation, followed by a leaky Rectified Linear Unit activation. The output values then become input values for the subsequent five residual blocks, containing only nine convolution layers while using a stride of 2.

A dropout layer drops half of the neurons to prevent overfitting. Then, the dropout layer's output is connected through the multi-head mechanism, with 8 heads on the 256-channel data, each head focusing on 32 dimensions. The multi-head's attention mechanism results are converted into a 256-dimensional vector. An adaptive max pooling layer reduces the dimensionality by downsampling each of the 256 feature maps to a single maximum value.

The NAS search space consists of a custom-made search space that combines convolutional operations and self-attention blocks. The initial kernel size controls the temporal receptive field by acting like a rectangular window function on the ECG signal, with smaller kernels capturing high-frequency features (i.e., IRBBB) and larger kernels longer temporal patterns (e.g., R-R interval for SB detection).

The dropout rate controls the regularisation strength by preventing overfitting. Low dropout rates are an advantage when the ECG has clear, consistent patterns, such as SB or NSR. High dropout rates are necessary for complex, morphological patterns, such as IRBBB.

The attention heads determine how many different ECG features can be analysed simultaneously. Fewer heads are good when a single pattern is dominant (e.g., consistent heart rate), while a larger number of heads improves multi-feature attention (e.g., QRS and T-wave analysis). The embedding dimension (256 channels) needs to be divisible by the number of heads so that each head can process an equal number of channels simultaneously.

We then conduct one NAS experiment with each dataset (baseline, rate, ventricular depolarisation and repolarisation), optimising the initial kernel size, dropout rate, and number of attention heads for each dataset. We subsequently retrain each of the four NN with the optimised parameters and test those on the PF12RED dataset.

2.3. Neural Architecture Search

A NAS design is composed of three major components: a search space, a search strategy and a performance estimation metric. The search space contains all possible network architectures. Our custom-made search space optimises the initial kernel size (possible options: 3, 9, 15, 21), number of attention heads (2, 4, 8, 16), and dropout rate (0.1, 0.3, 0.5, 0.7). A search strategy is then employed to identify the best-performing architectures within the chosen search space. We apply DARTS, a gradient-based approach, which transforms the discrete decision search space into a continuous, differentiable search space [11]. We choose DARTS because it is an established search strategy and a computational compromise between a RandomOneShot strategy (random selection from a supernet) and RegularisedEvolution, where all possible combinations are trained. Binary Cross-Entropy is used as the performance estimation metric.

With the three sets of determined parameters (initial kernel size, dropout rate, and attention heads), we train and validate four NN models (baseline, rate, depolarisation, and repolarisation). We then test the four NNs on the PF12RED arrhythmia dataset. NNs were trained for

50 epochs using a learning rate of 1×10^{-3}. The performance metrics we use are the area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve (AUPRC), and the F1-score.

3. Results

The initial kernel size is smallest for the baseline and ventricular depolarisation ($n = 3$), followed by ventricular repolarisation ($n = 9$) and rate-based arrhythmias ($n = 21$) NAS. The NAS results for the number of attention heads are 2 for all experiments but ventricular depolarisation ($n = 16$). The dropout rate is 0.5 for ventricular repolarisation and 0.7 for the rest.

All four experiments score an average AUROC score of over 0.5 as shown in table 1, with the baseline NN scoring the highest (0.62) and rate-based the lowest (0.55). For single experiments, the AUROC value is below 0.5 for TWI in the rate and ventricular repolarisation, and for SB in the ventricular depolarisation. The highest value is 0.76 for NSR in the ventricular depolarisation.

The averaged AUPRC in table 1 results were highest for ventricular depolarisation (0.64) and lowest for the rate-based experiment (0.57). Single AUPRC results were highest for TWI (between 0.98–0.99) and lowest for IRBBB (between 0.17–0.26).

The ventricular repolarisation experiment (see table 2) scores the highest (0.67) and ventricular depolarisation the lowest (0.54) averaged F1-score. We observe the lowest single score during the rate experiment for SB (0.10), and the highest during all experiments for TWI (0.99).

4. Discussion

The baseline and ventricular depolarisation NAS parameters result in the smallest available initial kernel size of 3. This parameter indicates that the NN needs to capture high-frequency features and short temporal patterns. We hypothesise that the low initial kernel size is due to the nature of the datasets: the baseline dataset contains over 30 different types of arrhythmias, while ventricular arrhythmias, such as IRBBB, are morphologically complex. For cardiac rhythms with more long-term dependencies, such as SB, the rate NAS search outputs the largest available kernel size of 21. The NAS for ventricular repolarisation results in 9, indicating a balanced complexity between depolarisation and rate-based cardiac rhythms.

All experiments output a dropout rate of 0.5 (ventricular repolarisation) or higher. The baseline, rate, and ventricular depolarisation all yield 0.7, suggesting strong regularisation to prevent overfitting due to the complexity of the signal.

For the baseline, rate, and ventricular repolarisation, the NAS determines that 2 is the optimal number of attention heads. The low number of attention heads is consistent when single patterns are dominant, such as SB (rate) or TWI (ventricular repolarisation). Although a higher attention head number could be expected for the baseline due to the arrhythmia heterogeneity, more than half (including NSR) of all rhythms are rate-dominant rather than morphology-dominant. The ventricular depolarisation NAS has the highest possible number of attention heads with 16, highlighting the morphological complexity behind rhythms like IRBBB.

The performance results demonstrate mixed evidence for the effectiveness of pathophysiology-informed NAS. Whilst the baseline achieves the highest overall AUROC of 0.62, individual rhythm-specific performance suggests some specialisation benefits. The rate-aware model, optimised with the largest kernel size (21) for temporal dependencies, shows reasonable NSR detection but performed poorly on SB classification (F1-score 0.10), implying that larger temporal windows may not adequately capture bradycardic patterns. The depolarisation-aware model, configured with maximum attention heads (16) and small kernels (3) for morphological complexity, achieves superior NSR performance (AUROC 0.76, AUPRC 0.83), supporting the hypothesis that increased attention mechanisms benefit complex rhythm detection. However, IRBBB classification proves challenging across all models (F1-scores 0.25–0.32), indicating that this morphologically complex arrhythmia requires further architectural refinement. TWI demonstrates optimal detection scores (0.98–0.99 AUPRC) across all experiments, although this likely reflected the high prevalence of TWI in the PF12RED test dataset ($98.1\,\%$) of ECGs rather than the actual model's performance.

A limitation is the multi-label nature of ECGs, i.e., having multiple arrhythmias diagnosed on one ECG. Other arrhythmias, apart from the targeted influence, affect the NAS training. For example, an ECG from the ventricular repolarisation pathophysiology-aware datasets had the diagnosis of TWI together with SB. We cannot conclude that ventricular repolarisation abnormalities solely influenced the parameter selection during ventricular repolarisation-aware NAS.

5. Conclusion

The NAS algorithm successfully identifies physiologically informed parameters that reflect the underlying pathophysiology of different cardiac rhythms. However, this specialisation does not translate into consistent performance improvements across all arrhythmia types during testing. Despite mixed performance results, these physiologically informed

Table 1: Area under the receiver operating characteristic curve (AUROC) and area under the precision recall curve (AUPRC) results for Baseline (BL), Rate (RT), and ventricular Depolarisation (VD) and Repolarisation (VR) experiments regarding Normal Sinus Rhythm (NSR), Sinus Bradycardia (SB), Incomplete Right Bundle Branch Block (IRBBB), and T-wave Inversion (TWI).

	NSR		SB		IRBBB		TWI		Average	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
BL	0.71	0.72	0.67	0.54	0.59	0.17	0.5	0.98	0.62	0.60
RT	0.55	0.64	0.55	0.42	0.63	0.23	0.46	0.98	0.55	0.57
VD	0.76	0.83	0.47	0.48	0.51	0.26	0.63	0.99	0.59	0.64
VR	0.74	0.81	0.60	0.45	0.63	0.19	0.44	0.98	0.6	0.61

Table 2: F1-score results for Baseline (BL), Rate (RT), and ventricular Depolarisation (VD) and Repolarisation (VR) experiments regarding Normal Sinus Rhythm (NSR), Sinus Bradycardia (SB), Incomplete Right Bundle Branch Block (IRBBB), and T-wave Inversion (TWI).

	NSR	SB	IRBBB	TWI	Average
BL	0.77	0.54	0.28	0.99	0.65
RT	0.77	0.55	0.32	0.99	0.66
VD	0.83	0.10	0.25	0.99	0.54
VR	0.80	0.56	0.34	0.99	0.67

parameters show promise for optimising diagnostic performance in specific cardiac rhythm classifications, representing a meaningful step toward domain-informed arrhythmia detection in sports cardiology. Future work could explore single-label datasets to isolate influences of different pathophysiologies and extend the framework to broader sports-related cardiac abnormalities.

References

[1] Walker J, Calkins H, Nazarian S. Evaluation of cardiac arrhythmia among athletes. The American Journal of Medicine December 2010;123(12):1075–1081. ISSN 00029343.

[2] La Gerche A, Wasfy MM, Brosnan MJ, Claessen G, Fatkin D, Heidbuchel H, Baggish AL, Kovacic JC. The athlete's heart—challenges and controversies. Journal of the American College of Cardiology October 2022; 80(14):1346–1362. ISSN 0735-1097.

[3] Huttin O, Selton-Suty C, Venner C, Vilain JB, Rochecongar P, Aliot E. Electrocardiographic patterns and long-term training-induced time changes in 2484 elite football players. Archives of Cardiovascular Diseases May 2018; 111(5):380–388. ISSN 1875-2136.

[4] Lampert R, Chung EH, Ackerman MJ, Arroyo AR, Darden D, Deo R, Dolan J, Etheridge SP, Gray BR, Harmon KG, James CA, Kim JH, Krahn AD, La Gerche A, Link MS, MacIntyre C, Mont L, Salerno JC, Shah MJ. 2024 hrs expert consensus statement on arrhythmias in the athlete: evaluation, treatment, and return to play. Heart Rhythm May 2024;ISSN 1547-5271.

[5] Fayyazifar N. An accurate cnn architecture for atrial fibrillation detection using neural architecture search. In 2020 28th European Signal Processing Conference (EUSIPCO). IEEE, January 2021; .

[6] Liu Z, Wang H, Gao Y, Shi S. Automatic attention learning using neural architecture search for detection of cardiac abnormality in 12-lead ecg. IEEE Transactions on Instrumentation and Measurement 2021;70:1–12. ISSN 1557-9662.

[7] Asadi M, Poursalim F, Loni M, Daneshtalab M, Sjödin M, Gharehbaghi A. Accurate detection of paroxysmal atrial fibrillation with certified-gan and neural architecture search. Scientific Reports July 2023;13(1). ISSN 2045-2322.

[8] Reyna M, Sadr N, Perez Alday E, Gu A, Shah A, Robichaux C, Rad A, Elola A, Seyedi S, Ansari S, Ghanbari H, Li Q, Sharma A, Clifford G. Will two do? varying dimensions in electrocardiography: the physionet/computing in cardiology challenge 2021.

[9] Munoz-Macho AA, Dominguez-Morales MJ, Sevillano-Ramos JL. An innovative 12-lead resting electrocardiogram dataset in professional football. Data in Brief June 2024; 54:110444. ISSN 2352-3409.

[10] Nejedly P, Ivora A, Smisek R, Viscor I, Koscova Z, Jurak P, Plesinger F. Classification of ecgs using ensemble of residual cnns with attention mechanism. In 2021 Computing in Cardiology (CinC). IEEE, September 2021; .

[11] Xiao Y, Qiu Y, Li X. A survey on one-shot neural architecture search. IOP Conference Series Materials Science and Engineering February 2020;750(1):012223. ISSN 1757-899X.

Address for correspondence:

Erik Vanegas Müller and Mauricio Villarroel
The Podium Institute, Old Road Campus Research Building,
University of Oxford, Oxford, OX3 7DQ, UK
erik.vanegasmuller@eng.ox.ac.uk
mauricio.villarroel@eng.ox.ac.uk