Athlete Hearts Confuse ECG Machines

Shaun Kyle [1]

[1]UNSW Graduate School of Biomedical Engineering

I. Background

Prolonged, intense exercise causes athlete hearts to undergo physiological adaptations. As a consequence, electrocardiogram (ECG) recordings of **athletes look abnormal** when compared to the **general population**.

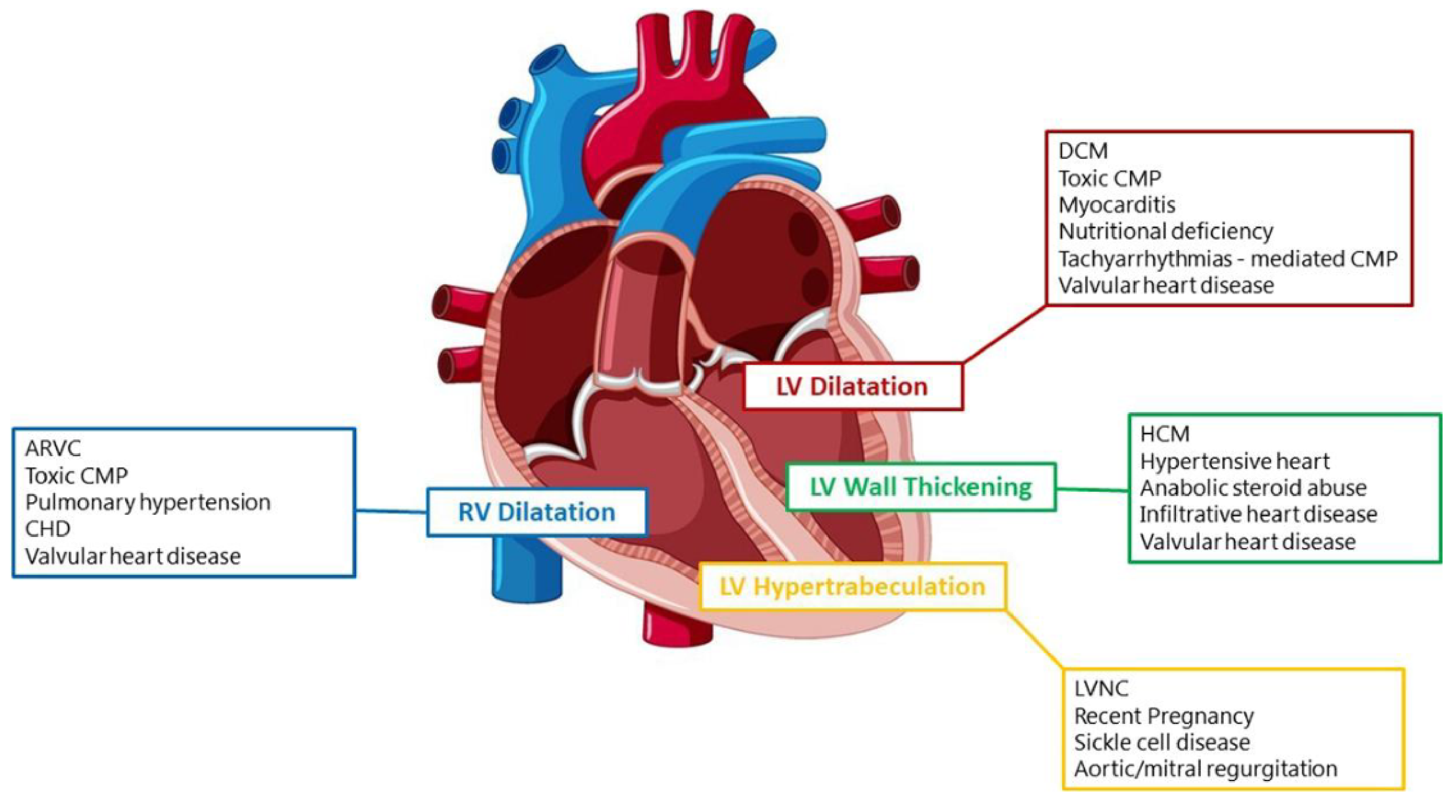


Figure 1. Potential misdiagnoses from athlete heart physiological adaptations

While cardiologists can be trained to interpret athlete ECG, many clinicians rely on the output of **ECG interpretation algorithms** to assist with clinical decision support. This could lead to misdiagnosis of athlete heart disease.

See the case study below for a real-world example of ECG misdiagnosis.

Case study: Machine from GE Healthcare

Between February and March 2020 at the University of Oslo, resting ECG recordings were collected from **28 elite athletes** across endurance sports such as rowing, kayaking, and cycling.

Each recording was interpreted by:

1. A cardiologist with specialization in athletes' hearts.
2. The Marquette 12SL algorithm on a MAC VUE 360 Electrocardiograph from GE Healthcare.

93% of recordings were determined to be "overall normal" by the cardiologist, while the machine determined that only 46% were normal.

The performance metrics below show misdiagnosis of specific findings.

Table 1. Performance of Marquette 12SL algorithm on endurance athletes

Finding	False-positives (%)	False-negatives (%)
Sinus bradycardia	61	0
Other sinus arrhythmia	10	0
Right bundle branch block	2	50

II. Aims

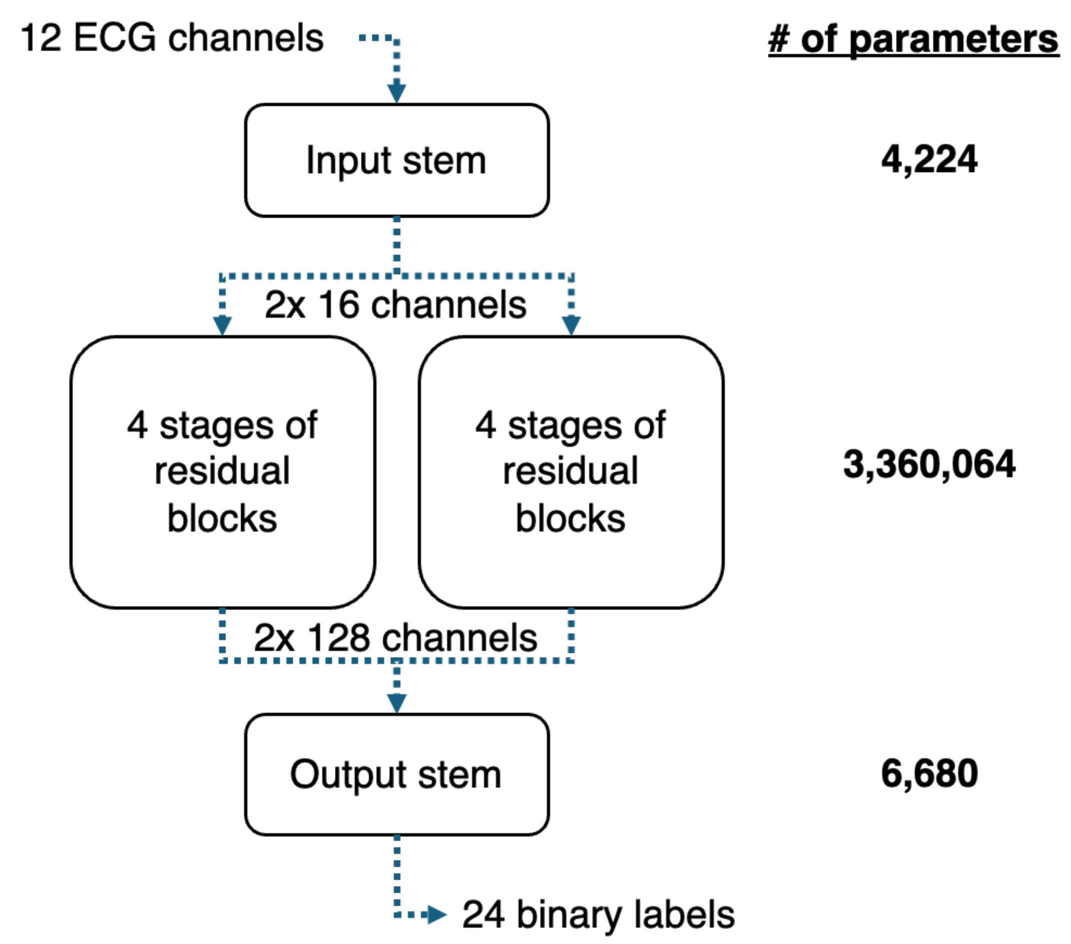
The problem of a source domain (general population) differing from a target domain (athletic population) is known as a **domain shift**.

Domain adaptation is a specific case of transfer learning where the task remains unchanged (e.g. diagnose Y using ECG recording from patient X), but the distribution of the target domain is shifted from the source domain.

1. Identify significant causes of this domain shift.
2. Devise a suitable domain adaptation method to adapt an ECG classifier trained on a general population to an athletic target population.

III. Methods

The ECG classifier chosen for this study is a deep neural network (DNN) developed at Seoul National University for the 2020 PhysioNet Challenge.



Approach 1: Finetuning on athletic data

One reason for reduced model performance is a **concept shift** for disease diagnosis criteria.

By comparing model predictions with actual cardiologist labels, we can fine-tune the output stem, which directly affects the classification output.

Approach 2: Reweighting source datasets

An athletic population has a far younger age distribution than general population. This is a **covariate shift** with age.

If we retrain the model from scratch, we can weight younger patient records with more importance in order to shift the effective training distribution closer to the target domain.

Types of domain shift

Domain shift can be defined mathematically as $P_S(X,Y) \neq P_T(X,Y)$ where X is input space (i.e. patients), Y is output space (i.e. diagnosis labels), and P_S/P_T are joint probability distributions of the source and target domains.

By decomposing $P(X,Y)$ into marginal (e.g. $P(X)$) and conditional (e.g. $P(Y|X)$) distributions, we can categorize causes of domain shift.

Covariate shift $P(X)$

Difference in **population demographics** between domains.

Table 2. Age summary statistics for source vs target populations

Patient population	Min	Max	Mean	Std
Norwegian athletes	20	42	**25**	4.7
Training (general population)	18	89	**60**	16.3

Label shift $P(Y)$

Prevalence of **diagnosis label** classes between domains

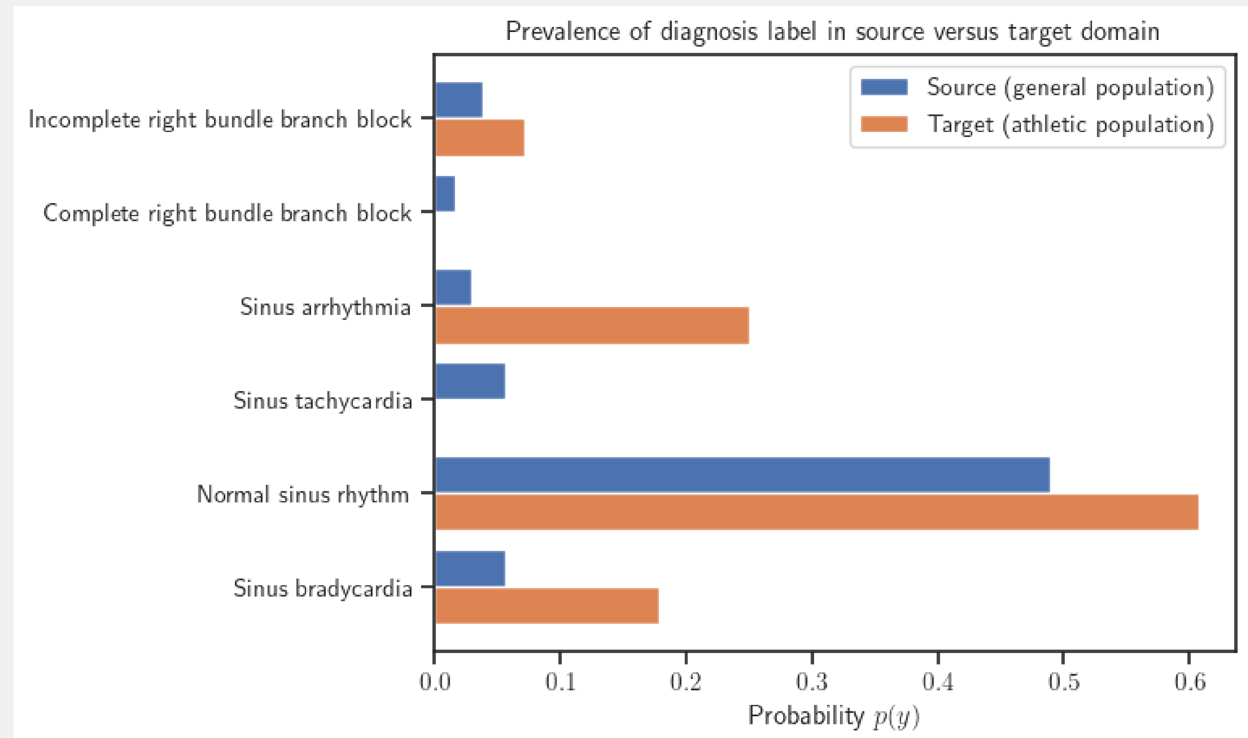


Figure 2. Label shift between $P_S(Y)$ and $P_T(Y)$

Concept shift $P(Y|X)$

Changing **criteria for diagnoses**. e.g. Threshold for sinus bradycardia diagnosis is under 40 bpm for athletes, 60 bpm otherwise.

IV. Key findings

Model	General domain (F1 score)	Athletic domain (F1 score)
Original	0.716	0.365
Adapted (Finetune)	0.802	0.470
Adapted (Reweight)	0.911	0.476