RECURRENT NEURAL NETWORKS FOR CLASSIFICATION OF HUMAN EMBRYONIC STEM CELL-DERIVED CARDIOMYOCYTES

by

Carolina Pacheco Oñate

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Master of Science.

Baltimore, Maryland

August, 2018

Abstract

Classification of human embryonic stem cell-derived cardiomyocytes (hESC-CMs) into phenotypes such as atrial-like or ventricular-like is important for applications in cardiac regenerative medicine and drug screening. However, a key challenge is the lack of ground truth labels for the phenotype of hESC-CMs: Whereas adult phenotypes are well-characterized in terms of the shape of their action potentials (APs), the understanding of how the shape of the AP of immature CMs relates to that of adult CMs remains limited. Recently, a new metamorphosis distance has been proposed to determine if a query immature AP is closer to a particular adult AP phenotype. However, the metamorphosis distance is difficult to compute making it unsuitable for classifying a large number of CMs.

This thesis proposes two recurrent neural networks (RNNs) with long short-term memory (LSTM) units for classifying hESC-CMs. The first network is trained using a semi-supervised approach, in which the parameters of the network are learned by minimizing a loss function consisting of two terms: a su-

pervised term that uses labeled data obtained from computational models of adult CMs, and an unsupervised term that uses a contrastive loss to encourage the labels of similar APs (as measured by the metamorphosis distance) to be the same. The second network is trained using a domain adaptation approach that captures the domain shift between immature and adult cells by adding a term to the loss function that penalizes their maximum mean discrepancy (MMD) in feature space.

Experiments confirm the benefit of integrating information from both adult and stem cell-derived domains in the learning scheme and show that the proposed semi-supervised method generates results similar to the state of the art (94.73%) with clear computational advantages when applied to new samples. Experimental results on the domain adapted learning approach confirm that it not only is more computational efficient but also outperforms the state of the art in terms of clustering quality.

In summary, the main contributions of this thesis are to formulate the classification of hESC-CM APs in the framework of artificial neural networks and to show that this new formulation improves with respect to the state of the art for this task in terms of both performance and computational efficiency.

Thesis Committee

Dr. René Vidal (advisor)

Professor, Department of Biomedical Engineering

Johns Hopkins University

Dr. Leslie Tung

Professor, Department of Biomedical Engineering

Johns Hopkins University

Dr. Archana Venkataraman

Assistant Professor, Department of Electrical and Computer Engineering

Johns Hopkins University

Acknowledgments

I would like to thank my advisor, Dr. René Vidal, for believing in me and giving me the opportunity to be part of his lab. His continuous guidance and support were fundamental for the development of this thesis. I would also like to thank Dr. Leslie Tung and Dr. Archana Venkataraman for their insightful comments.

I want to thank my labmates at the "Computer Vision, Dynamics and Learning Lab" for being always willing to help. It is wonderfully challenging to be surrounded by such a group of smart and hardworking people. In particular I would like to thank Dr. Giann Gorospe, because his experience in the classification of action potentials and his understanding of the problematic made my first steps in this topic much smoother.

I want to thank my family and friends for their never-failing support, I would not be the person that I am if I had not learned from you all. Especially I would like to thank my parents, Maritza and Héctor, who have always been my biggest models.

ACKNOWLEDGMENTS

Finally, I would like to thank CONICYT BECAS CHILE 73170418 for funding part of this work.

Contents

CONTENTS

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Ischaemic heart disease (IHD), which refers to the group of diseases characterized by deficient supply of blood to the heart, is the leading cause of death not only in the US but also globally, taking the lives of more than 9.4 million people around the world in 2016 [2]. Being an important public health issue, worldwide efforts are focused on prevention, in terms of promoting healthy lifestyles to reduce risk factors such as unhealthy diet or physical inactivity [3]. Although advances in pharmacological treatments and revascularization surgeries have significantly reduced IHD mortality over the past decades, it remains the leading cause of death, which has motivated the search of innovative sources of treatment such as cell therapy [4].

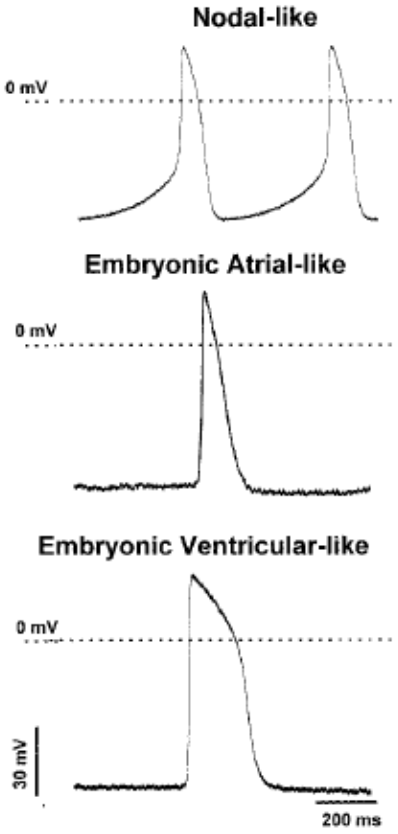
Cell-based therapies for IHD generally rely on paracrine mechanisms of action to promote the formation of new blood vessels or attenuate adverse ventricular remodeling, but it is too late for this type of treatment in acute myocardial infarction (MI) survivors [5], for whom the presence of scar tissue often leads to heart failure due to the limited self-regenerative capacity of the adult human heart. Cardiomyocyte transplantation then arises as an interesting alternative treatment to promote direct remuscularization in post-MI heart failure. Experimental work in the field demonstrates the feasibility and favorable results of cardiomyocyte transplantation in the remuscularization of myocardial scar tissue in animals [6–8], however determining the most appropriate mechanism of delivery and finding a large-scale source of cardiomyocytes (CMs) are just some of the important challenges to overcome before cardiomyocyte transplantation becomes a common practice at a clinical level.

The fact that human embryonic stem cells (hESCs) can functionally and structurally differentiate into CMs was experimentally confirmed in 2001 [9], and since then important efforts have been put into improving the efficiency and control of cardiac differentiation protocols. Nowadays, their unquestioned cardiomyogenic potential [9,10] and the well-established protocols for their isolation and maintenance [11] have positioned hESCs as one of the most promising sources of CMs not only for cell-based cardiac repair [12], but also for other applications such as drug screening [13, 14]. The goal of in vitro drug screen-

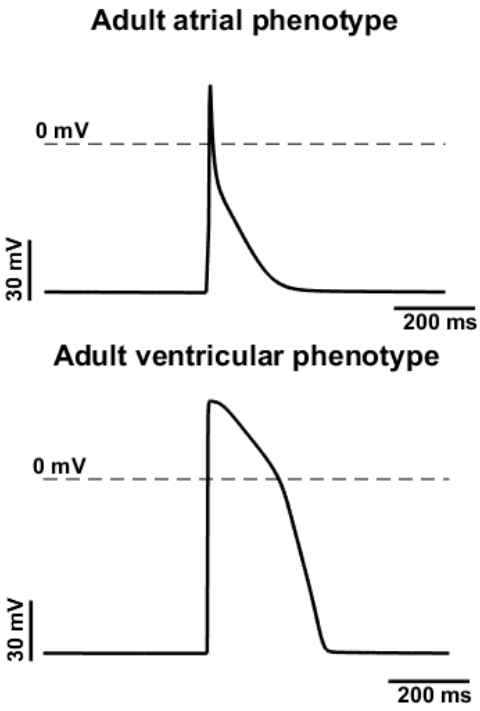
ing is to study the effect that pharmacological compounds have in cell cultures in order to assess their potential efficacy or toxicity in a target tissue. This has been traditionally done using animal cells, but these animal models do not always have been capable of predicting the pharmacological response later observed in humans. In this context, the response of differentiated hESCs to pharmacological compounds could potentially provide better indication on how the target tissue would react, since their human embryonic origin removes interspecies variability [15, 16].

The cardiac nature of human embryonic stem cell-derived cardiomyocytes (hESC-CMs) is confirmed not only by cellular ultrastructure and the expression of cardiac-specific genes, but also by extracellular electrical activity [9, 17]. However, while different studies show that the electrophysiological characteristics of hESC-CMs approach those of human embryonic CMs through differentiation, there are important electrophysiological differences between hESC-CMs and adult CMs [16–18] (See Figure 1.1). This fact rises concerns regarding the use of hESC-CMs in regenerative medicine, because even if the hESC-CMs are well-integrated into the myocardium, "they could still induce rhythm disturbances if their single-cell electrophysiologic properties are sufficiently different from those of the adult host myocardium" [12]. Regarding drug screening, the underdeveloped nature of hESC-CMs is also a source of concern since their immature state can confound their response, impacting the effect of the

pharmacological stimuli under study.



(a) Major hESC-CM AP subtypes described by He et. al (source: [17]).



(b) Example of typical adult atrial and ventricular AP phenotypes. They correspond to atrial Nygren model [19] and ventricular O'Hara-Rudy model [20] with nominal parameters.

Figure 1.1: Comparison between hESC-CM and adult APs phenotypes. Embryonic atrial-like and ventricular-like phenotypes present slower upstroke velocity and depolarized resting membrane potential compared to typical adult phenotypes.

While hESC-CMs are developmentally immature by definition, as described above, applications like regenerative medicine and drug screening require them to be understood in the context of adult cells that have been under development for years. Therefore, although the potential impact of hESC-CMs has

been broadly noticed, their use in clinical applications is still hampered by the limited understanding of the properties of hESC-CMs and how they relate to adult CMs. In this context, a first step to build this understanding is to study the presence of the major adult CM phenotypes (atrial, ventricular, nodal) in hESC-CM populations.

The characterization of hESC-CMs can be addressed by means of the expression of specific genes or ion channel-encoding subunits [18, 21], but there are some interesting reasons to study their electrophysiological features instead. First of all, action potentials (APs) constitute the net balance between ion currents in the cell membrane, and therefore they provide functional information by summarizing complex interactions between the cell and its environment in just one signal, conveniently reducing this problem to classification of single temporal series. Second, it is well-known that the shape of the APs is clearly distinguishable between different adult CM phenotypes. This led to the development of different computational models that describe the electrophysiology of atrial CMs [19, 22], ventricular CMs [20, 23, 24] and, to a lesser degree, nodal/pacemakers CMs [25]. Therefore it is reasonable to argue that if hESC-CMs were to exhibit adult-like phenotypes, their APs would reflect it. Lastly, high resolution imaging techniques for mapping the electrophysiology of cells in vitro [26], as opposed to patch-clamp methods, now provide the opportunity to get hundreds of recordings from precise locations without mechanical

disruption of the cardiac cell aggregates [27], making considerable amounts of hESC-CM electrophysiological data available to better characterize them.

1.2 Fundamental challenges

Identifying the presence of adult CM phenotypes in hESC-CM populations is a challenging problem from multiple perspectives. One of them, and perhaps the most fundamental for designing a classifier, is that there is no consensus about the existence of such phenotypes in hESC-CMs (this is also the case for human induced pluripotent stem cell-derived cardiomyocytes iPSC-CMs [28–31]). Therefore, most of the hESC-CM data is unlabeled, and the small amount of labeled data available has not been categorized based on well-established criteria.

The limited and unreliable nature of ground truth-labeled hESC-CM data is problematic for two main reasons. First, the performance of classic machine learning classification approaches strongly depends on the amount and quality of the labeled data available for training. One potential workaround would be the use of unsupervised clustering methods, since they do not require labels to group similar samples together. However in this case hESC-CM samples must be related to adult CM phenotypes, so simply identifying different groups in the hESC-CM population is not enough. Second, comparison of different

classification methods as well as tuning of their parameters require a metric of performance, which usually is the classification accuracy that they achieve in labeled datasets for test and validation, respectively. However, the scarcity of ground truth labels for hESC-CMs forces the use of alternative and indirect metrics, such as clustering quality criteria, to evaluate the performance of the classification algorithms.

In this context, computational models of adult CM electrophysiology [19,20, 22–25] constitute a convenient unlimited source of labeled data, since different samples can be generated by randomizing the value of some of their parameters [32]. This synthesized labeled data of adult CMs then can be used to train and validate classification algorithms to be tested afterwards in hESC-CM populations. Nevertheless, traditional machine learning algorithms assume that training and testing datasets are sampled from the same probability distribution [33], an assumption that does not hold in this case since it has been already established that hESC-CM APs resemble embryonic but not adult CM APs [16,18]. Thus, the use of computational models to overcome the scarcity of labeled hESC-CM data is undoubtedly appealing, but it also opens new challenges in terms of how to leverage adult CM information appropriately when the main task must be performed on data coming from a different probability distribution, i.e., lying in a different domain.

Beyond the lack of labeled data and the challenges that arise when adult

CM data come into play, there are additional constraints regarding computational complexity. One of them arises from the fact that the space of action potentials is not Euclidean, for example, the Euclidean interpolation between two APs does not always produce a signal that resembles an AP [34]. Therefore, important work has been done using shape theory to find a metric that better describes the space of APs. This metric is called metamorphosis, it was introduced in [35], first applied to cardiac APs in [34], and corresponds to the minimum energy required to interpolate between two APs when this interpolation is modeled as a diffeomorphism acting on an evolving template. The computation of this metric involves solving an optimization problem in two variables (the infinitesimal change in deformation and the evolution of the template). Therefore, it takes 12 times more to compute the metamorphosis distance than it takes to compute the Euclidean distance, even when the most efficient metamorphosis computation available is used [1]. Thus, if a classification method requires computing distances between thousands APs, the decision of whether to use the Euclidean or metamorphosis distance involves an important tradeoff between accuracy and computational time.

1.3 Prior work

Both classification and characterization of APs in adult CMs have been traditionally assessed by looking at some basic AP features. Some of the most commonly used ones are depicted in Figure 1.2 and correspond to: (i) the action potential duration APD_p, which is the time that it takes to reach $p\%$ of the repolarization; (ii) the resting membrane potential RMP; (iii) the maximum upstroke velocity $\max\left[\frac{dV}{dt}\right]$; (iv) the maximum diastolic membrane potential MDP; and (v) the action potential amplitude APA.

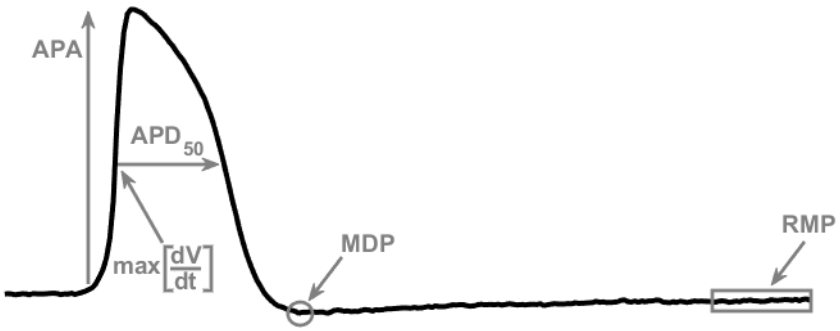


Figure 1.2: Basic action potential features.

In 2003 He et al. [17] for the first time studied the existence of different types of hESC-CMs by looking at their APs. Based on knowledge about adult CM phenotypes and simple AP features (APD_{50}, APD_{90}, $\max\left[\frac{dV}{dt}\right]$, MDP, APA), they identified 3 hESC-CM subtypes: nodal-like, embryonic atrial-like, and embryonic ventricular-like. The last two subtypes were named embryonic to emphasize that they exhibit slow upstroke velocity $\max\left[\frac{dV}{dt}\right]$ and depolarized RMP compared to adult CMs, resembling embryonic CMs instead (See Figure

1.1).

After this initial study other groups have tried to further characterize the variability of hESC-CM APs respect to different factors, such as their in vitro maturation [18], their response to pharmacological stimuli [16], and their spatial organization [21]. None of these studies had as primary goal to design a standard classification method; all of them rather classify the samples based on ad-hoc thresholds adjusted to their experimental data. For example, Sartiani et al. [18] suggests the existence of atrial-like and ventricular-like subtypes using a threshold of 200 ms in APD_{70}, Peng et al. [16] on the other hand indicates that a threshold of 150 ms in APD_{90} suitably separates atrial/pacemaker APs from ventricular subtypes, while Vestergaard et al. [21] uses the ratio $\frac{APD_{20}}{APD_{70}}$ and the upstroke velocity $\max\left[\frac{dV}{dt}\right]$ (with thresholds 0.35 and $25\frac{V}{s}$ respectively), to distinguish between nodal-like, atrial-like, and ventricular-like subtypes.

The data gathered in the studies mentioned above confirms the heterogeneity of hESC-CM populations, and their methodology of analysis demonstrates the need of categorizing hESC-CM data in terms of adult CM phenotypes. However, the use of handcrafted features and subjective criteria for their classification not only makes it difficult to translate between datasets, but also discards most of the information contained in the signals. Their transferability is restricted even further by the limited amount of samples analyzed in each study

(between 12 and 128 APs)[1], making any conclusion drawn from these small datasets unlikely to be representative of a larger population. Moreover, the limited understanding about these AP features in hESC-CM APs and their tenuously justified selection make these ad-hoc classification methods further questionable.

Fortunately, the improvement of high-throughput in vitro recording techniques has enabled the generation of large datasets, which consequently allows the use of machine learning and signal analysis methods to delve into this problem. In 2014 Gorospe et al. [36] applied for the first time an automatic clustering method to a large hESC-CM dataset (6940 APs) using the entire AP as a feature. They used a spectral clustering algorithm based on Euclidean similarities between APs to study the number of clusters observed in a population of hESC-CMs. They found that their dataset is better described by two clusters than by three or more, and although the algorithm is not informed by adult CM phenotypes, visual inspection of the average AP per cluster suggests the presence of atrial-like and ventricular-like subtypes. Zhu et al. [27] showed later that this automatic clustering method provides better clustering results than automatic methods based only on basic AP features, confirming that information relevant for the task is lost when only one or two AP features are selected.

[1]The quality of the methods to induce cell differentiation and the available techniques to acquire electrophysiological data at that time limited the size of the datasets.

The algorithm presented in [36] successfully proved the benefit of using the whole AP signal for this task, as opposed to the traditional approach of considering only basic AP features. However, as any clustering algorithm it is intrinsically limited since it does not associate a label to the clusters it creates. To address this issue, a method that can capture the relationship between hESC-CM and adult CM phenotypes is needed.

In this context, in [34] Gorospe et al. took a different approach proposing a 1 Nearest Neighbor (1NN) method to classify hESC-CM APs. In this algorithm 20 synthetic adult APs generated from computational models of atrial [19] and ventricular [20] phenotypes were used as templates. For any given hESC-CM AP, metamorphosis distances to each one of the 20 templates were computed, and then the label of the closest template was assigned to it (atrial-like or ventricular-like). This method leverages the adult CM data by simulating the maturation process of the AP. They showed that metamorphosis distance works better than Euclidean distance for this purpose not only because the interpolants lie in AP space, but also because it leads to better results in terms of classification accuracy in a small dataset (52 APs). However, the high computational cost of the metamorphosis distance made it unfeasible to apply this 1NN classifier to a larger dataset. In 2015, Gorospe et al. [1] proposed a much faster way to compute the metamorphosis distance by finding a closed form solution to part of the optimization problem. They applied the 1NN classifier

to a large dataset (6940 APs) and showed that the use of the metamorphosis distance leads to better results than the use of the Euclidean distance in terms of clustering quality.

However, the main drawback of this approach is that although the metamorphosis distances are computed efficiently, they are not used in an efficient way: even if many APs have been classified before, 20 metamorphosis distances need to be computed every time a new sample needs to be analyzed. In [1] they reported that it took 13 hours to classify 6940 APs with the most efficient implementation, which would double if we want to use 40 templates instead of 20, or if we want to classify twice as large a dataset.

In summary, while multiple studies confirm the heterogeneity of hESC-CM data, traditional classification methods based on basic AP features discard relevant information and are hardly transferable. The use of machine learning and signal analysis techniques to classify large datasets has not only provided automatic tools that can be applied to different datasets, but also established the advantages of studying the whole AP signal. However, as progress on high-throughput recording techniques continues, larger datasets will be available, and existing algorithms will not be able to scale up as needed due to their computational complexity.

1.4 Thesis contributions

This thesis addresses the problem of classifying hESC-CM APs according to adult CM phenotypes, for which we propose recurrent neural networks (RNNs) with long short-term memory (LSTM) units as classifiers. Unlike the state-of-the-art 1NN classifier with metamorphosis distances proposed in [1], neural networks are models with learning capabilities, which is advantageous for scalability purposes. The task we are addressing corresponds to classification of time series; thus neural networks with feedback connections, named recurrent neural networks, are the appropriate framework because they are designed to learn time-varying patterns using the whole time series as input. In particular we decided to use RNNs with LSTM units [37] because of their great performance in applications to speech recognition [38] and activity recognition [39], among others.

LSTM units are recurrent blocks with a particular architecture, especially designed to overcome the problem of vanishing/exploding gradient, common in the training of recurrent neural networks. The key idea behind LSTM units is the use of multiplicative gate variables that learn to open and close access to the error flow across the layers in a way that depends on the task they are trained for [40]. This uniqueness has provided LSTM networks with a distinctive wide range of memory capacity, leading them to be successfully applied

to many disciplines. Explicitly related to the cardiology field, they have been used to automate the analysis of electrocardiogram (ECG) records, both from a time series classification approach [41–43] and from a predictive approach for anomaly detection [44, 45]. However, these methods train the LSTM networks using fully annotated datasets, which is not applicable to classification of hESC-CM APs because of the lack of labels. Moreover, to the best of our knowledge, no RNNs of any type have been used to classify hESC-CMs.

Therefore, our first contribution is to propose an RNN-based classifier of hESC-CM APs that not only has the potential to reduce the computational cost of the classification task, since complex relationships can be learned by the network during training, but also opens the opportunity to apply to this problem many tools that are constantly emerging in the fast-growing field of artificial neural networks. However, training the proposed classifier is not trivial due to the limited availability of ground truth labels.

In general, training methods are classified as supervised, unsupervised or semi-supervised depending on how they use ground truth-labeled data. While supervised methods use ground truth labels for training, unsupervised methods only have access to unlabeled data, so they make use of secondary information such as similarity between samples instead. Semi-supervised methods lie in between both approaches, using labeled as well as unlabeled samples for training. Semi-supervised methods have drawn attention in many applications

in which data acquisition is a relatively fast and inexpensive procedure compared to ground truth labeling; hence there are large datasets available from which only a small subset of samples is labeled.

In this context, our second contribution is to propose a semi-supervised approach for training the proposed classifier, which overcomes the scarcity of labeled hESC-CM data by exploiting the abundance of labeled adult CM data that can be obtained via simulation of electrophysiological models for the typical adult phenotypes. Our semi-supervised approach uses a novel loss function that combines a crossentropy loss for adult APs (supervised part) and a contrastive loss for hESC-CM APs (unsupervised part). We evaluated this algorithm in a 6940 hESC-CM dataset and showed it is a more efficient way to use similarities between APs: it significantly outperforms the 1NN scheme in terms of clustering quality when only Euclidean distances are available, and when metamorphosis distance is applied it generates similar to state-of-the-art results with significantly less distance computations.

The main assumption of the semi-supervised approach is that a single network can simultaneously classify both adult and hESC-CM APs. However, the semi-supervised approach does not take into account the fact that the underlying probability distributions of adult and hESC-CM APs are different, which might be useful information to train the network. Currently there is an emerging field in machine learning called domain adaptation, which specifically stud-

ies cases of mismatch between training (source) and test (target) domains, aiming at building classifiers trained in labeled data from the source domain that will perform well in the target domain [33].

Our third contribution is to propose a domain adaptation approach to train the LSTM classifier, in which the probability distributions of adult CMs and hESC-CMs are forced to be similar in a latent space. This similarity is imposed by minimization of the Maximum Mean Discrepancy (MMD) [46] in the feature space corresponding to the output of a hidden layer in the network architecture. This approach also assumes that a single network can simultaneously solve the classification task in both domains but, unlike the semi-supervised approach, it performs an adaptation step to take into account the domain mismatch. When tested in a 6940 hESC-CM dataset it outperforms both the semi-supervised approach and the state-of-the-art results in terms of clustering quality. Moreover, the use of a parametric model for the classifier allows reduction of the time that it takes to evaluate the whole large dataset from 13 hours (state of the art) to just a couple of seconds. The proposed classifier also proves to have great transferability capacity since it reaches the same classification accuracy as the state-of-the-art method in a small and completely different dataset without retraining, which further demonstrates the computational advantages of the proposed method.

1.5 Thesis outline

The remainder of the thesis is organized as follows. In Chapter 2 a semi-supervised learning approach to classification of hESC-CM APs using RNNs is presented. Its performance is experimentally evaluated and compared to a baseline supervised approach and to the state-of-the-art method (1NN classifier with metamorphosis distances presented in [1]). In Chapter 3 an unsupervised domain adaptation approach to classification of hESC-CM APs is presented and also integrated with the semi-supervised approach proposed in Chapter 2. Experimental results are presented to evaluate its performance. Finally, main conclusions and future work are discussed in Chapter 4.

Chapter 2

A semi-supervised approach to classification of APs

2.1 Problem formulation

Let $\Omega_e = \{\mathbf{x}_j^e\}_{j=1}^{N_e}$ be an *unlabeled* hESC-CM APs dataset, hereafter referred to as embryonic, where the sequence $\mathbf{x}_j^e = \{x_j^e(k) \in \mathbb{R}\}_{k=1}^{K}$ represents the jth embryonic AP and K is the total number of samples in one cycle length. Let $\Omega_a = \{(\mathbf{x}_i^a, y_i^a)\}_{i=1}^{N_a}$ be a *labeled* adult dataset, where $\mathbf{x}_i^a = \{x_i^a(k) \in \mathbb{R}\}_{k=1}^{K}$ is the ith adult AP and $y_i^a \in \{0, 1\}$ is its ground truth label ($y_i^a = 0$ denotes atrial and $y_i^a = 1$ denotes ventricular). We consider the problem of assigning a label $\hat{y}_j^e$ to each $\mathbf{x}_j^e \in \Omega_e$, where $\hat{y}_j^e = 0$ denotes atrial-like and $\hat{y}_j^e = 1$ denotes ventricular-like.

2.2 A nearest-neighbor approach to classification of APs

In this section we review the approach presented by Gorospe et al. [1], which uses a 1 nearest-neighbor (1NN) classifier: it assigns to the embryonic AP $\mathbf{x}_j^e$ the label of its closest adult sample as described by

$$\hat{y}_j^e = \{y_{i^*}^a \mid i^* = \underset{i \in \{1,2,\cdots,N_a\}}{\operatorname{argmin}} \; d\left(\mathbf{x}_j^e, \mathbf{x}_i^a\right)\}, \tag{2.1}$$

where $d\left(\cdot,\cdot\right)$ is a function that measures the distance between two APs.

There are many ways in which the distance between two time series can be defined, but only Euclidean and metamorphosis distances have been used in the classification of APs, so they are described below.

2.2.1 Euclidean distance

The Euclidean distance d_E between two APs is defined as the Euclidean norm of the vector formed by the difference between both sequences:

$$d_E^2(\mathbf{x}_j^e, \mathbf{x}_i^a) = \frac{1}{\sigma_M^2} \left\|\mathbf{x}_j^e - \mathbf{x}_i^a\right\|_2^2 = \frac{1}{\sigma_M^2} \sum_{k=1}^{K} \left(x_j^e(k) - x_i^a(k)\right)^2. \tag{2.2}$$

In (2.2) σ_M is a normalization parameter introduced to make the values of

the Euclidean distance comparable to the values of other distances. Using the Euclidean distance, the interpolation between $\mathbf{x}^e$ and $\mathbf{x}^a$ considering S interpolation steps is described by

$$x(k,s) = x^e(k) + s\frac{(x^a(k) - x^e(k))}{S} \text{ with } s \in \{0, 1, \cdots, S\}, \tag{2.3}$$

where the interpolants $\mathbf{x}(s) = \{x(k,s) \in \mathbb{R}\}_{k=1}^{K}$ are such that $\mathbf{x}(0) = \mathbf{x}^e$ corresponds to the initial embryonic AP, and $\mathbf{x}(S) = \mathbf{x}^a$ is the target adult AP.

Figure 2.1a depicts an example of this point-wise Euclidean interpolation between an embryonic AP and a ventricular adult AP for the case where $S = 4$. As it can be seen, the initial $(s = 0)$ and final $(s = 4)$ shapes correspond to APs, however the interpolants ($s = 1, 2, 3$) do not look like physiological APs due to a dimple introduced by the interpolation method. This example shows that the Euclidean distance, although simple to compute, is not adequate for modeling and classification of APs, because Euclidean interpolation leads to intermediate APs that do not lie in the space of APs. In other words, the space of APs is not Euclidean, and therefore another metric is needed to compare APs.

2.2.2 Metamorphosis distance

Instead of performing a point-wise comparison of the two APs, the metamorphosis distance finds an interpolation path between the two APs with minimum deformation. The interpolation is modeled as a diffeomorphism acting on an evolving template and the amount of deformation is measured by defining a suitable norm on the diffeomorphism and the template.

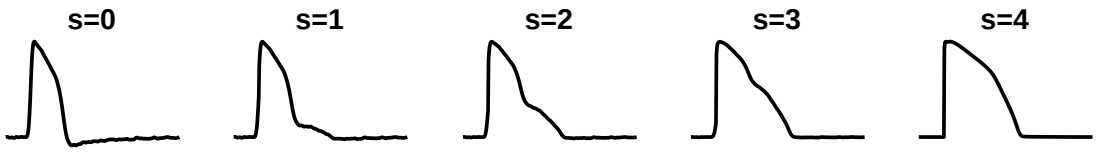
More specifically, consider again the case in which the interpolation between an embryonic AP $\mathbf{x}^e$ and an adult AP $\mathbf{x}^a$ is performed in S interpolation steps. The metamorphosis path $x(k,s)$ must not only meet the boundary constraints $x(k,0) = x^e(k)$ and $x(k,S) = x^a(k)$, but also minimize the energy

$$d_M^2(\mathbf{x}^e, \mathbf{x}^a) = \min_{\mathbf{x},\mathbf{v}} \sum_{s=0}^{S-1} \|v(k,s)\|_{V_d}^2 + \tfrac{1}{\sigma_M^2}\|x(k+v(k,s),s+1) - x(k,s)\|_2^2, \qquad (2.4)$$

which depends on: (i) the infinitesimal change in deformation v that primarily takes care of the adjustment in the temporal domain; and (ii) the evolution of the template $x(k,s)$ that accounts for changes of amplitude. σ_M is a balancing parameter between both terms and $\|\cdot\|_{V_d}^2$ is a discretized Sobolev norm, i.e. $\|v\|_{V_d}^2 = \langle L_d v, L_d v\rangle$, where L_d is a discrete approximation of a Sobolev operator L. For additional details please refer to [1, 34].

The metamorphosis distance d_M is defined as the squared root of the optimal energy in (2.4). Figure 2.1b shows the interpolation generated by the

metamorphosis method between an embryonic AP ($s = 0$) and a ventricular adult AP ($s = 4$). Unlike Euclidean interpolation, in this method not only the initial and final shapes but also the interpolants ($s = 1, 2, 3$) resemble physiological APs, which is indicative of a better exploration of the space of APs.



(a) Euclidean interpolation

(b) Metamorphosis interpolation

Figure 2.1: Example of interpolation between an embryonic AP ($s = 0$) and an adult ventricular AP ($s = 4$) using (a) Euclidean and (b) metamorphosis distances.

Results presented in [1] show that the metamorphosis distance is more appropriate than the Euclidean distance to study the similarity between APs not only because of the shape of the interpolants, but also because it generates better clustering quality in a population of hESC-CMs. However, computing the metamorphosis distance is expensive since it requires to solve an optimization problem with respect to x and v, which significantly limits its use.

2.3 A semi-supervised LSTM approach to classification of APs

The main drawback of the 1NN classifier (2.1) is that N_a distance computations are required every time a new sample $\mathbf{x}_j^e$ needs to be classified. Although this might not be a problem in some applications, it is critical when classifying APs because the cost of each metamorphosis computation is far from being negligible. We hypothesize that the use of a parametric classifier would overcome this drawback, since the computational cost of classifying new samples is reduced at the expense of a training stage where the optimal parameters are learned.

2.3.1 An RNN with LSTM units as a classifier

We propose a recurrent neural network (RNN) with long short-term memory (LSTM) units as classifier. RNNs are neural networks with feedback connections, which have been traditionally used to discover time-varying patterns in time series data. However, in their simple form they are hard to train by backpropagation because feedback information usually gets diminished as the length of the input increases, and therefore they perform poorly in discovering patterns with long-time delays.

In order to overcome this problem a new type of recurrent units, named LSTMs, was introduced by Hochreiter and Schmidhuber in 1997 [37]. These new recurrent blocks preserve feedback information at the same time that use multiplicative gate units to learn, depending on the task, when it is relevant to propagate the feedback information. LSTMs have recently gained popularity for time-series classification due to their great performance in challenging tasks such as activity recognition and speech recognition, and therefore their application to classification of APs is promising.

After the original LSTM architecture was proposed in 1997, several variants have emerged presenting different advantages and disadvantages for specific applications (a detailed study on these variants can be found in [47]). The classifier we propose uses the LSTM architecture presented by Gers et al. [48], which is depicted in Figure 2.2 and described by the following set of equations:

$$\begin{aligned} i\left(k\right) &= \sigma\left(W_i x\left(k\right) + U_i h(k-1) + b_i\right) \in \mathbb{R}^p \\ f\left(k\right) &= \sigma\left(W_f x\left(k\right) + U_f h\left(k-1\right) + b_f\right) \in \mathbb{R}^p \\ o\left(k\right) &= \sigma\left(W_o x\left(k\right) + U_o h\left(k-1\right) + b_o\right) \in \mathbb{R}^p \\ c\left(k\right) &= f\left(k\right) \circ c\left(k-1\right) + i\left(k\right) \circ \tanh\left(W_c x\left(k\right) + U_c h\left(k-1\right) + b_c\right) \in \mathbb{R}^p \\ h\left(k\right) &= o\left(k\right) \circ \tanh\left(c\left(k\right)\right) \in \mathbb{R}^p, \end{aligned} \tag{2.5}$$

where its key elements are the input gates $i(k)$, forget gates $f(k)$ and output gates $o(k)$ that modulate the evolution of its state $c(k)$ and output $h(k)$ accord-

ing to the input sequence $x(k) \in \mathbb{R}^m$. The scalar p denotes the dimension of the LSTM layer, $W_i, W_f, W_o, W_c \in \mathbb{R}^{p \times m}$ correspond to the input weight matrices, $U_i, U_f, U_o, U_c \in \mathbb{R}^{p \times p}$ are the recurrent weight matrices, $b_i, b_f, b_o, b_c \in \mathbb{R}^p$ are the bias vectors, $\circ$ denotes Hadamard product and $\sigma(z) = \dfrac{1}{1+e^{-z}}$ corresponds to the point-wise logistic sigmoid function.

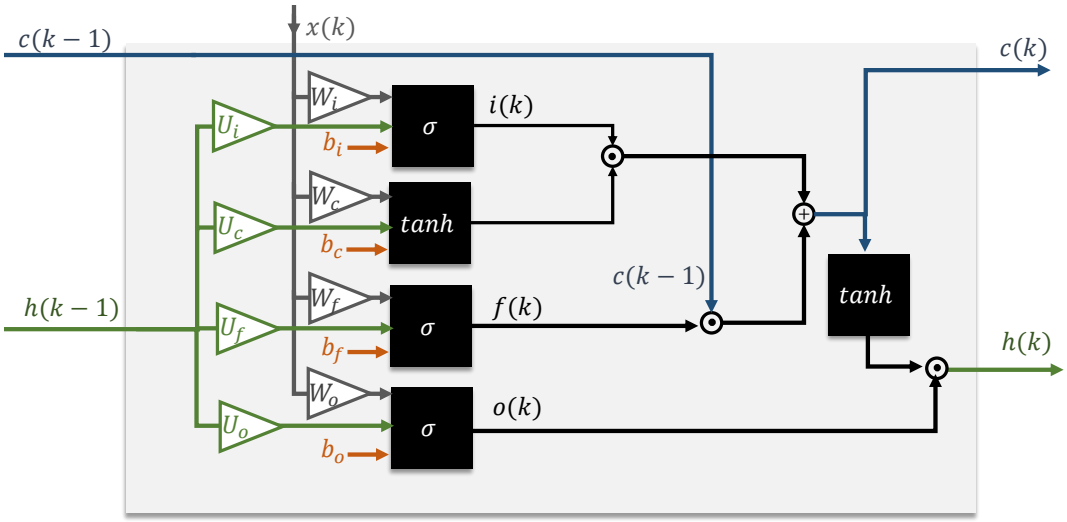


Figure 2.2: LSTM block.

The architecture of the proposed classifier is depicted in Figure 2.3 and consists of one input layer, one hidden LSTM layer of dimension $p = 3$, and a single sigmoid unit as the output layer. This sigmoid unit operates only in the last value of the hidden layer output, once all the input sequence $\mathbf{x} = \{x(k) \in \mathbb{R}\}_{k=1}^{K}$ has been processed by the LSTM layer. The predicted label is given by $\hat{y} = \sigma\left(h(K)^T W + b\right)$, where $W \in \mathbb{R}^3$ and $b \in \mathbb{R}$ are parameters of the output unit; therefore the hidden recurrent LSTM layer can be simply seen as a feature

extraction network and $h(K) \in \mathbb{R}^3$ as the corresponding feature vector.

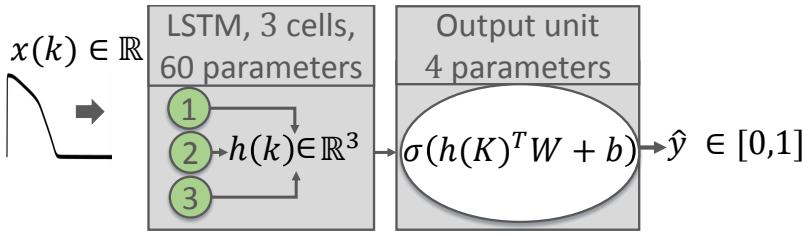


Figure 2.3: Proposed semi-supervised classification approach.

To simplify notation, let $\mathcal{W} = \{W_i, W_f, W_o, W_c, U_i, U_f, U_o, U_c, b_i, b_f, b_o, b_c, W, b\}$ be the set of parameters of the proposed classifier. For a given set of parameters $\mathcal{W}$, we will represent the classifier as the function $f_{\mathcal{W}}(\mathrm{x}) = \hat{y}$ that maps an action potential x to a predicted label $\hat{y}$.

The dimension of the proposed classifier, determined by the number of hidden layers and their dimension, was chosen to be small to avoid overfitting. Thus, the total number of parameters of the network is 64 (60 parameters in the hidden layer, and 4 parameters in the output layer).

It should be noted that this architecture can be easily extended to address more complex cases, for example, increasing the number of units in the output layer allows for multi-class problems. Definitions and derivations presented in the following sections can also be easily extended to more complex scenarios.

2.3.2 Proposed semi-supervised loss function

Along with defining the classifier architecture, it is necessary to define a loss function to guide the search of the optimal parameters for the task. As commonly done in classification problems, we use the binary crossentropy loss defined by

$$\ell(y, \hat{y}) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y}) \tag{2.6}$$

to quantify how close the LSTM prediction $\hat{y}$ is to the actual label y. More specifically, given N_a adult APs $\{\mathbf{x}_i^a\}_{i=1}^{N_a}$ and their labels $\{y_i^a\}_{i=1}^{N_a}$, our supervised loss corresponds to the average of the individual losses:

$$\frac{1}{N_a} \sum_{i=1}^{N_a} \{-y_i^a \log(\hat{y}_i^a) - (1 - y_i^a) \log(1 - \hat{y}_i^a)\}. \tag{2.7}$$

Now, while we do not have labels for the embryonic APs $\{\mathbf{x}_j^e\}_{j=1}^{N_e}$, we can still use $\ell(\hat{y}_j^e, \hat{y}_{j'}^e)$ to compare the predicted labels for two different embryonic APs. Intuitively, we would like similar APs to have the same labels, and dissimilar APs to have different labels. Therefore, we define the contrastive unsupervised loss as follows

$$\ell_u\left(\hat{y}_j^e, \hat{y}_{j'}^e\right) = s_{(j,j')} \cdot \ell\left(\hat{y}_j^e, \hat{y}_{j'}^e\right) + \left(1 - s_{(j,j')}\right) \cdot \ell\left((1 - \hat{y}_j^e), \hat{y}_{j'}^e\right), \tag{2.8}$$

where $s_{(j,j')} \in [0, 1]$ represents the similarity between AP $\mathbf{x}_j^e$ and AP $\mathbf{x}_{j'}^e$, such

that $s_{(j,j')} = 1$ if the APs are identical. We define the similarity between two APs based on their distance $d\left(\mathbf{x}_j^e, \mathbf{x}_{j'}^e\right)$ (Euclidean or metamorphosis) as follows

$$s_{(j,j')} = \exp\left(-\frac{d^4\left(\mathbf{x}_j^e, \mathbf{x}_{j'}^e\right)}{\sigma_s^4}\right), \tag{2.9}$$

where the design parameter σ_s is chosen as $\sigma_s^4 = \overline{d^4}$, i.e. its fourth power is the average of the fourth power of the pair-wise distances between embryonic APs.

The unsupervised loss function presented in (2.8) penalizes prediction disagreements when samples are similar and prediction agreements when samples are dissimilar. One may wonder if both terms are really needed or if we may be able to use only the first term. The reason for needing both terms is that weighting the loss according to the similarity between the inputs (first term in the summation) can lead to the pathological case in which all the samples are assigned to the same label. Therefore, the second term in the summation is required to counterbalance.

The basic concept behind contrastive loss functions consists of applying a specific loss if the pair of samples being evaluated are similar, and a different loss if they are dissimilar. It was first presented in [49] to discover a low dimensional representation of the data but, like other works that have used it later [50, 51], they assume the availability of binary similarity factors given by prior knowledge, labels or a clustering algorithm, and they use loss func-

tions that are based on Euclidean distances between the outputs of the network. Therefore, our approach is novel in the sense that we consider similarity factors as a continuous variable, so we do not assume prior knowledge about clusters in the data.

Given N_e embryonic APs, we average the unsupervised loss (2.8) over all the possible pairs of samples as follows

$$\frac{1}{N_e\left(N_e-1\right)}\sum_{j=1}^{N_e}\sum_{j'\neq j}\ell_u\left(\hat{y}_j^e,\hat{y}_{j'}^e\right). \tag{2.10}$$

Integrating supervised and unsupervised terms of the loss, we obtain the semi-supervised loss function

$$\frac{1-\lambda}{N_a}\left(\sum_{i=1}^{N_a}\ell\left(y_i^a,\hat{y}_i^a\right)\right)+\frac{\lambda}{N_e\left(N_e-1\right)}\left(\sum_{j=1}^{N_e}\sum_{j'\neq j}\ell_u\left(\hat{y}_j^e,\hat{y}_{j'}^e\right)\right), \tag{2.11}$$

where $\lambda\in[0,1]$ is a balancing parameter between supervised and unsupervised parts.

In this way, the proposed semi-supervised loss function leverages labeled adult CM data as well as unlabeled embryonic CM data for training the classifier.

2.4 Metrics to evaluate classification and clustering

2.4.1 Classification accuracy

When ground truth labels are available, the most direct way to evaluate the performance of a binary classification algorithm is to compare the predicted labels $\hat{y}_i \in \{0, 1\}$ to the ground truth labels $y_i \in \{0, 1\}$ in a test dataset. More specifically, let $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ be a labeled test dataset composed of N samples and let $\{\hat{y}_i\}_{i=1}^N$ be the corresponding labels predicted by a given classification algorithm. The classification accuracy corresponds to the number of correctly classified samples over the total number of samples

$$Acc\left(\{y_i\}_{i=1}^N, \{\hat{y}_i\}_{i=1}^N\right) = 1 - \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \,. \tag{2.12}$$

In the case of the recurrent neural network, the output of the classifier is continuous $\hat{y}_i = f_{\mathcal{W}}(\mathbf{x}_i) \in [0, 1]$; therefore there is a thresholding process before computing the classification accuracy. The predicted label is considered to be 1 if $\hat{y}_i \geq 0.5$ and 0 otherwise.

2.4.2 Davies-Bouldin Index (DBI)

Since ground truth labels are rarely available for embryonic APs, the Davies-Bouldin Index (DBI) [52] is considered as a measure of clustering quality. Let $\Omega_0 = \left\{\mathbf{x}_j^e \mid \hat{y}_j^e < 0.5\right\}$ and $\Omega_1 = \left\{\mathbf{x}_j^e \mid \hat{y}_j^e \geq 0.5\right\}$ be the sets containing the different clusters. Let μ_0 and μ_1 be the average signals per cluster given by

$$\mu_0 = \frac{1}{|\Omega_0|} \sum_{\mathbf{x}_j^e \in \Omega_0} \mathbf{x}_j^e \text{ and } \mu_1 = \frac{1}{|\Omega_1|} \sum_{\mathbf{x}_j^e \in \Omega_1} \mathbf{x}_j^e. \tag{2.13}$$

Let S_0 be the mean distance from elements of Ω_0 to μ_0, and similarly for S_1 as follows:

$$S_0 = \frac{1}{|\Omega_0|} \sum_{\mathbf{x}_j^e \in \Omega_0} d\left(\mathbf{x}_j^e, \mu_0\right) \text{ and } S_1 = \frac{1}{|\Omega_1|} \sum_{\mathbf{x}_j^e \in \Omega_1} d\left(\mathbf{x}_j^e, \mu_1\right). \tag{2.14}$$

Let $M_{01} = d\left(\mathcal{U}_0, \mathcal{U}_1\right)$ be the distance between the averages of the clusters. The DBI is defined as the ratio between the intra-cluster dispersion and the distance between clusters

$$DBI\left(\Omega_0, \Omega_1\right) = \frac{S_0 + S_1}{M_{01}}, \tag{2.15}$$

and should be as small as possible. For computational reasons, and since the Euclidean distance d_E is a good approximation of the metamorphosis distance d_M for small distances, the intra-cluster dispersions S_0 and S_1 are computed

using d_E, whereas the distance between clusters M_{01} is computed using d_M.

2.4.3 Variation of Information (VI)

In order to evaluate the convergence of the clustering results generated by the network, we use the Variation of Information (VI) [53] as a metric of distance between partitions. Only the case in which the dataset is divided into two classes is presented, since that corresponds to the problem we are studying.

Let $\Omega_e = \{\mathbf{x}_j^e\}_{j=1}^{N_e}$ be a dataset. A clustering $\mathcal{C}$ divides Ω_e into mutually disjoint subsets Ω_0 and Ω_1. Formally,

$$\mathcal{C} = \{\Omega_0, \Omega_1\} \text{ such that } \Omega_0 \cap \Omega_1 = \emptyset \text{ and } \Omega_0 \cup \Omega_1 = \Omega_e. \tag{2.16}$$

Consider the case of randomly picking a sample from the dataset. If we assume that each sample has equal probability of being picked, the probability that the outcome belongs to cluster Ω_k is given by $P\{k\} = \dfrac{|\Omega_k|}{|\Omega_0| + |\Omega_1|}$. The uncertainty about the cluster to which the random sample would belong is described by

$$H(\mathcal{C}) = -\sum_{k=0}^{1} P\{k\} \log\left(P\{k\}\right), \tag{2.17}$$

which is called "*entropy associated to clustering* $\mathcal{C}$". $H(\mathcal{C})$ is non-negative and takes value 0 only when there is not uncertainty (all the samples belong to the same cluster).

Let $\mathcal{C}^* = \{\Omega_0^*, \Omega_1^*\}$ be a different clustering of dataset Ω_e. The probability that a sample simultaneously belongs to cluster Ω_k in clustering $\mathcal{C}$ and to cluster $\Omega_{k^*}^*$ in clustering $\mathcal{C}^*$ is given by

$$P\left\{k, k^*\right\} = \frac{|\Omega_k \cap \Omega_{k^*}^*|}{|\Omega_0| + |\Omega_1|}. \tag{2.18}$$

If we pick a random sample from Ω_e and we want to know to which cluster it belongs in clustering $\mathcal{C}$, the "*mutual information between clusterings $\mathcal{C}$ and $\mathcal{C}^*$*" given by

$$I(\mathcal{C}, \mathcal{C}^*) = \sum_{k=0}^{1} \sum_{k^*=0}^{1} P\{k, k^*\} \log\left(\frac{P\{k, k^*\}}{P\{k\}P\{k^*\}}\right) \tag{2.19}$$

represents how much uncertainty is reduced in average if we know to which cluster the sample belongs in clustering $\mathcal{C}^*$ (the other way around is also true). Then, the conditional entropy $H(\mathcal{C}|\mathcal{C}^*) = H(\mathcal{C}) - I(C, C^*)$ represents how much uncertainty is left in $\mathcal{C}$ when $\mathcal{C}^*$ is given. The VI between the two clusterings is defined as the sum of their conditional entropies:

$$VI\left(\mathcal{C}, \mathcal{C}^*\right) = H(\mathcal{C}|\mathcal{C}^*) + H(\mathcal{C}^*|\mathcal{C}) = H(\mathcal{C}) + H(\mathcal{C}^*) - 2I(\mathcal{C}, \mathcal{C}^*), \tag{2.20}$$

and therefore it is a way to measure how different they are. VI satisfies the metric axioms: non-negativity, symmetry and triangle inequality [53]. Also, if

$\mathcal{C}$ and $\mathcal{C}^*$ have at most K clusters with $K \leq \sqrt{n}$, where n is the total number of samples,

$$VI(\mathcal{C}, \mathcal{C}^*) \leq 2\log(K). \tag{2.21}$$

In our case the number of clusters is $K = 2$; therefore we define the normalized VI as

$$\overline{VI}(\mathcal{C}, \mathcal{C}^*) = \frac{H(\mathcal{C}) + H(\mathcal{C}^*) - 2I(\mathcal{C}, \mathcal{C}^*)}{2\log(2)} \in [0, 1]\,. \tag{2.22}$$

2.5 Experiments

2.5.1 Adult CM data

The O'Hara-Rudy model (ORd) [20] is used to generate synthetic examples of ventricular adult CM APs. Its main parameters are maximum conductances and permeabilities, and their nominal values are shown in Table 2.1. One thousand random sets of parameters were independently drawn from uniform distributions between 80% and 120% of their nominal values.

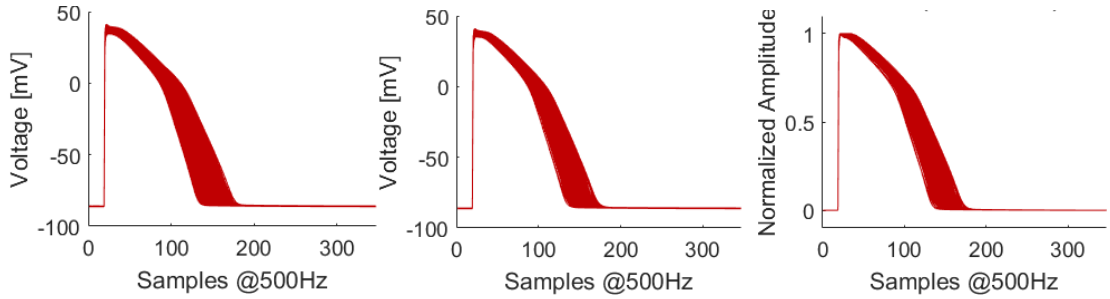
The ORd model was paced at 1.5Hz and run for 100 beats for each set of parameters. The effect of each given set of parameters is considered to be captured by the last AP of the simulation. Figure 2.4a shows the population of 1000 synthetic ventricular APs generated by this method, to which the Sparse

Name	**Symbol**	**Nominal value**	**Unit**
Max conductance, Na^+	$\overline{G_{Na}}$	75	$mS/\mu F$
Max conductance, transient outward K^+	$\overline{G_{to}}$	0.02	$mS/\mu F$
Max conductance, rapid delayed rect. K^+	$\overline{G_{Kr}}$	0.046	$mS/\mu F$
Max conductance, slow delayed rect. K^+	$\overline{G_{Ks}}$	0.0034	$mS/\mu F$
Max conductance, inward rect. K^+	$\overline{G_{K1}}$	0.1908	$mS/\mu F$
Max conductance, Na^+/Ca^{2+} exchange	$\overline{G_{NaCa}}$	0.0008	$\mu A/\mu F$
Max conductance, background K^+	$\overline{G_{Kb}}$	0.003	$mS/\mu F$
Max conductance, sarcolemmal Ca^{2+} pump	$\overline{G_{pCa}}$	0.0005	$mS/\mu F$
Max conductance, late Na^+ current	$\overline{G_{Na\ell}}$	0.0075	$mS/\mu F$
Permeability to Na^+/K^+ ATPase current	P_{NaK}	30	cm/s
Permeability to Ca^{2+} background current	P_{Cab}	$2.5e^{-8}$	cm/s
Permeability to Na^+ background current	P_{Nab}	$3.75e^{-10}$	cm/s
Permeability to Ca^{2+} current	P_{Ca}	0.0001	cm/s

Table 2.1: Ventricular model parameters.

Modeling for Representatives Selection (SMRS) method [54] was applied to select the subset of 150 APs shown in Figure 2.4b. As it can be noted, the subset of 150 APs seems to appropriately cover the heterogeneity of the 1000 APs dataset, even when the number of samples is significantly smaller. Normalization was applied to each AP so that its maximum voltage is 1 and its resting membrane potential is 0, as depicted in Figure 2.4c.

A similar process was followed using the Nygren model [19] to generate synthetic examples of atrial adult CM APs. In the original formulation, the Nygren model was tuned to achieve long-term stability with its nominal parameters (shown in Table 2.2) at a pacing rate of 1Hz, for which Nygren et al. incorporated an external Na^+ influx to their model. This causes a lack of charge conservation as well as the absence of true mathematical steady-states in the model, which have been studied [55] and even corrected in later mod-



(a) 1000 APs (b) 150 selected APs (c) 150 selected APs (normalized)

Figure 2.4: Ventricular adult CM data: (a) 1000 synthetic examples generated by ORd model [20], (b) subset of 150 representatives selected by SMRS method [54], and (c) 150 representatives after normalization.

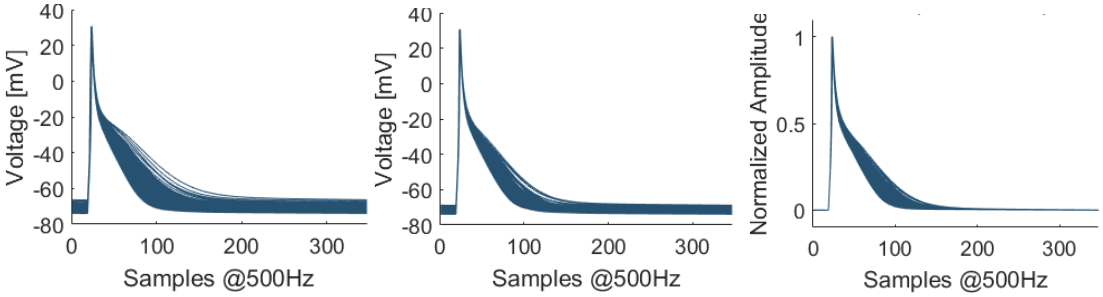
els [22]. However, we decided to use this model because it is the one that has been used in previous classifications of hESC-CMs [1,34], and also because the normalization process helps reduce the effect of unstable simulations, which in general are associated with higher resting membrane potential (RMP). In order to avoid sets of parameters that would increase the RMP too far from its physiological range, the values associated with $\overline{G_{K1}}$ and $\overline{G_{Cab}}$ were coupled to ensure a reasonable balance between inward rectifier K^+ current and Ca^+ background current, keeping fixed the ratio observed between their nominal values. Thus, for the six parameters presented in Table 2.2 only five independent random variables were used.

The Nygren model was paced at 1.5Hz and run for 100 beats for each of the 1000 random sets of parameters generated. Figure 2.5a shows the 1000 synthetic examples corresponding to the last beat of each simulation, Figure

Name	Symbol	Nominal value	Unit
Max conductance, L-Type Ca^{2+} current	$\overline{G_{CaL}}$	6.75	nS
Max conductance, slow delayed rect. K^+	$\overline{G_{Ks}}$	1	nS
Max conductance, rapid delayed rect. K^+	$\overline{G_{Kr}}$	0.5	nS
Max conductance, inward rect. K^+	$\overline{G_{K1}}$	3	nS
Max conductance, Na^+ background current	$\overline{G_{Nab}}$	0.060599	nS
Max conductance, Ca^{2+} background current	$\overline{G_{Cab}}$	0.078681	nS

Table 2.2: Atrial model parameters.

2.5b shows the subset of 150 representatives selected by the SMRS method, and Figure 2.5c shows their normalized version.



(a) 1000 APs (b) 150 selected APs (c) 150 selected APs (normalized)

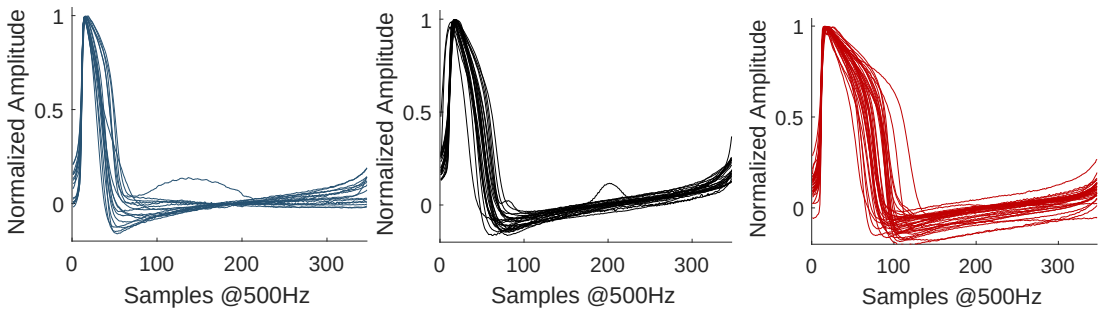
Figure 2.5: Atrial adult CM data: (a) 1000 synthetic examples generated by the Nygren model [19], (b) subset of 150 representatives selected by SMRS method [54], and (c) 150 representatives after normalization.

2.5.2 hESC-CM data

2.5.2.1 Single cell recording data

The dataset generated by [17] contains 16 atrial embryonic-like, 24 nodal-like, and 36 ventricular embryonic-like hESC-CM APs, manually labeled based

on basic AP features. These single cell recordings were obtained under spontaneous beating of the cells; therefore each signal has a different length. We use the algorithm presented by [56] to adjust them to a pacing rate of 1.5Hz, and we normalized them to have maximum amplitude of 1, as shown in Figure 2.6.



(a) 16 atrial-like APs (b) 24 nodal-like APs (c) 36 ventricular-like APs

Figure 2.6: Single cell hESC-CM recordings [17]: (a) atrial embryonic-like, (b) nodal-like, and (c) ventricular embryonic-like.

This dataset is small, and the quality of its labels is questionable due to the lack of well-known criteria. However, it is highly valuable because it is the only source of labeled hESC-CM APs available to us. Therefore, it is not used for training or validation, but only for testing the performance of the classifiers.

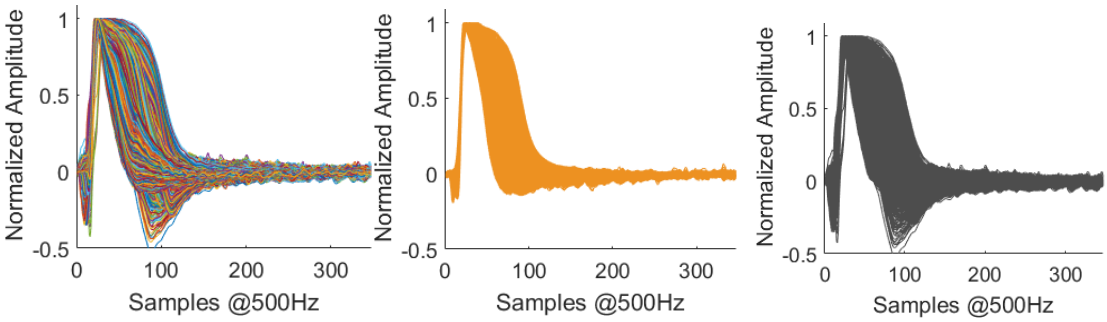
2.5.2.2 Optical mapping data

The dataset from [27] corresponds to 6940 hESC-CM APs recordings obtained by Dr. Renjun Zhu in the Cardiac Bioelectric Systems Laboratory of Johns Hopkins University from 9 cell aggregates paced at 1.5Hz and optically mapped at a sampling rate of 500Hz. Figure 2.7a depicts the 9 cell aggregates,

where each one of the pixels corresponds to a recording site. The APs were averaged over beating cycles, and a 5×5 boxcar spatial filter was applied for denoising. Each AP was normalized so that its maximum voltage is 1 and its resting membrane potential is 0, as it can be seen in Figure 2.7b.

Only 1600 APs coming from 2 cell aggregates were used for training and validation (see Figure 2.7c), but labels were predicted for the whole dataset.

(a) Cell aggregates



(b) Complete dataset (6940 APs)

(c) Training and validation subset (1600 APs)

(d) Test subset (5340 APs)

Figure 2.7: Optical mapping dataset [27] (yellow indicates training and validation data, gray indicates test data): (a) 9 cell aggregates, (b) Complete APs dataset, (c) Training and validation data, and (d) Test data.

2.5.3 Implementation details

The classifier architecture was implemented in Keras [57] with TensorFlow backend and trained using the RMSProp optimizer (with initial learning rate $\epsilon = 0.003$). The network was initialized according to the default methods: states of the LSTM layer were initialized orthogonally, the forget bias was set as $b_f = 1$ [48], and the rest of the weights were initialized by the method of Glorot and Bengio [58].

The 150 representatives of ventricular APs and the 150 representatives of atrial APs formed a dataset of $N_a = 300$ adult samples. On the other hand, the training and validation subset of the optical mapped hESC-CMs formed a dataset of $N_e = 1600$ embryonic samples. A set of 100 random mini-batches, each one formed of 19 samples ($n_a = 3$ adults and $n_e = 16$ embryonic), was built at the beginning of training[1]. Therefore, the gradient of the supervised part of the loss function is estimated by the average of the n_a observations at each iteration. Similarities between consecutive embryonic samples in each mini-batch were computed, generating $n_e - 1$ observations to estimate the gradient of the unsupervised part of the loss function at each iteration.

90 mini-batches were used for training, and 10 for validation. Figure 2.8

[1]The use of a finite predefined set of mini-batches does not correspond to the classic definition of mini-batch stochastic gradient descent (mini-batch SGD), in which a random mini-batch is built at every iteration of the algorithm. However, the computational cost of the metamorphosis algorithm makes the computation of distances between different APs at every iteration impractical. We conjecture that this modification does not significantly affect the guarantees of mini-batch SGD, and we verify it experimentally in Section 2.5.4.4.

illustrates the training scheme described above. One epoch is considered to be a complete pass of the training dataset, which in our case corresponds to 90 iterations of the optimization algorithm.

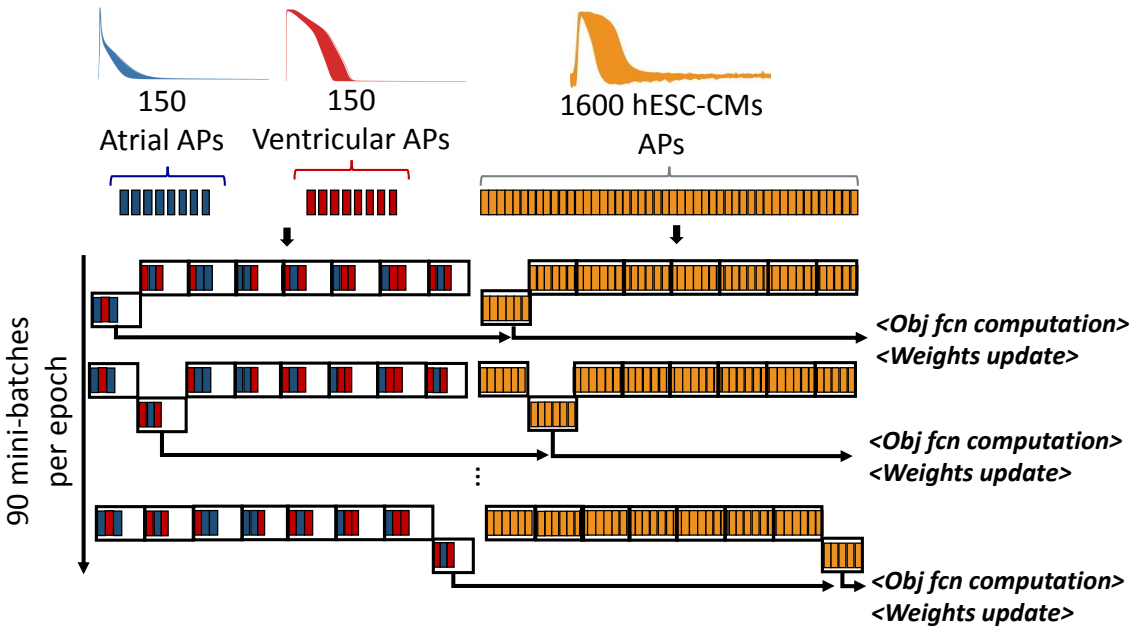


Figure 2.8: Training scheme for the semi-supervised learning approach.

Three cases are studied: Supervised learning $\lambda = 0$ (Sup-LSTM), Semi-supervised learning $\lambda = 0.1$ with Euclidean distances (Semi-LSTM-E), and Semi-supervised learning $\lambda = 0.1$ with metamorphosis distances (Semi-LSTM-M).

In each case the network was trained 5 times with the same initialization for the weights (100 epochs for the Sup-LSTM network and 200 epochs for the Semi-LSTM networks). Each one of these 5 runs of the optimization algorithm is referred to as a "trial". The variability observed across trials for

a given method (Sup-LSTM, Semi-LSTM-E or Semi-LSTM-M) is attributable to the stochastic implementation of the optimization algorithm by Keras with TensorFlow backend. The metamorphosis parameter was set as $\sigma_M = 0.3$.

2.5.4 Results

In this section the results of the Sup-LSTM, Semi-LSTM-E, and Semi-LSTM-M networks are presented individually. A comparative analysis of their performances with respect to the performance of the state of the art (1NN classification with metamorphosis distances) is presented in Section 2.5.5.

Due to the high computational cost of the metamorphosis algorithm, we implemented a modified version of mini-batch stochastic gradient descent (SGD). Experimental results regarding the effect that this modification has on the behavior of mini-batch SGD are presented in Section 2.5.4.4.

2.5.4.1 Sup-LSTM

Figure 2.9 depicts the training results for the Sup-LSTM case, i.e. $\lambda = 0$. Each color represents a different trial, and whenever they are indistinguishable it is because they have converged to the same value. The first column shows the loss function evaluated in the training set (top) and validation set (bottom). Although the training and validation losses oscillate during train-

ing[2], they both converge close to zero for all trials.

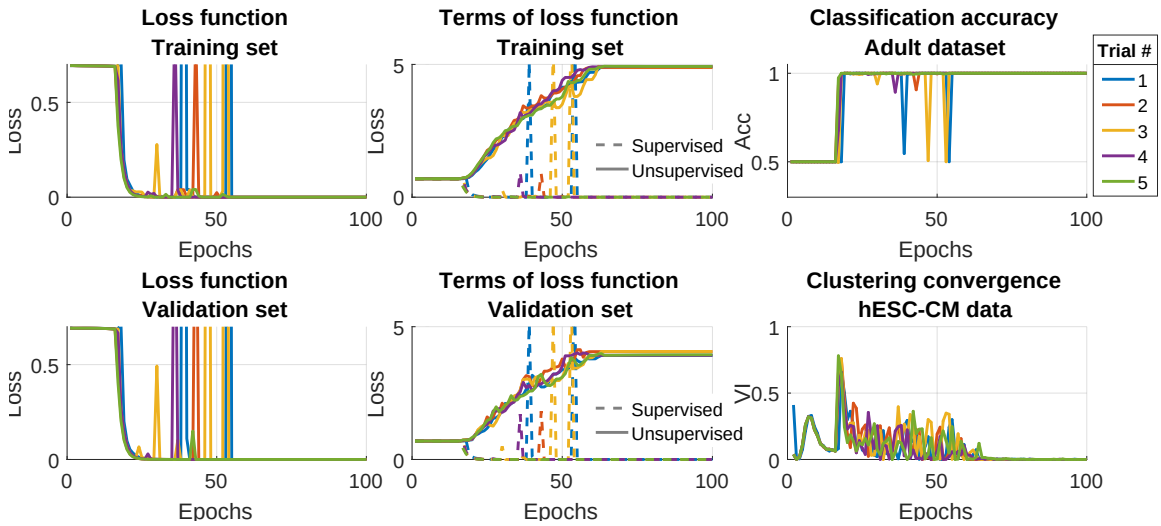


Figure 2.9: Training results of the Sup-LSTM network.

The second column of Figure 2.9 shows both the supervised and unsupervised parts[3] of the loss function evaluated on the training (top) and validation sets (bottom). As it can be seen, the supervised part converges to a small value in both cases while the unsupervised part converges to a large value in both cases, which makes sense since the latter is not being optimized ($\lambda = 0$).

The third column of Figure 2.9 depicts the classification accuracy evaluated in the 300 adult CM APs (top), and the normalized variation of information of embryonic clusterings computed between consecutive epochs (bottom). They show that at each trial the network learns to correctly classify the samples in

[2]It is unclear why all trials present spikes around epoch #50, it might have to do with the specific shape of the objective function in parameter space.

[3]In this case the unsupervised term was computed using similarity factors based on metamorphosis distances, but the same behavior is observed if Euclidean distances are applied instead

the adult domain (classification accuracy close to 1), while the clustering result converges (VI close to 0).

However, while each trial converges to a particular clustering result, they do not converge to the exact same clustering. Figure 2.10 shows the sets of network weights at the last epoch of training for each one of the five trials. As it can be seen, their sets of parameters, although similar, are not exactly the same at the last epoch of training[4].

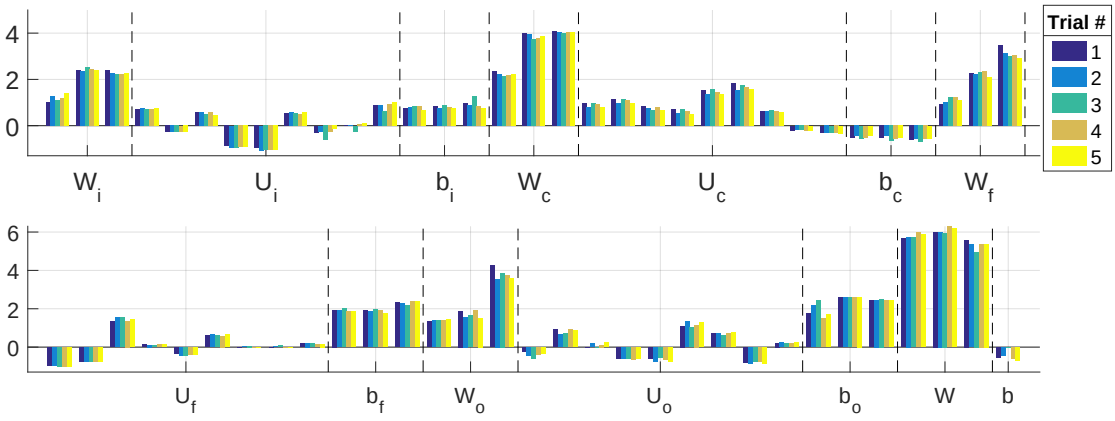


Figure 2.10: Weights of the Sup-LSTM network at the last epoch of training for five different trials.

The output of the network $\hat{y}$ is interpreted as the probability that a given sample belongs to the ventricular-like class (as opposed to belonging to the atrial-like class). Therefore, if the output of the network for a given sample is $\hat{y} \geq 0.5$ it is classified as ventricular-like, and if $\hat{y} < 0.5$ it is classified as atrial-like. We would like to combine the outputs of the different trials leveraging the

[4]Different sets of parameters can generate the same clustering results in a overparameterized model. Therefore, the fact that the trials converge to different sets of parameters does not fully explain the difference in clustering observed across trials.

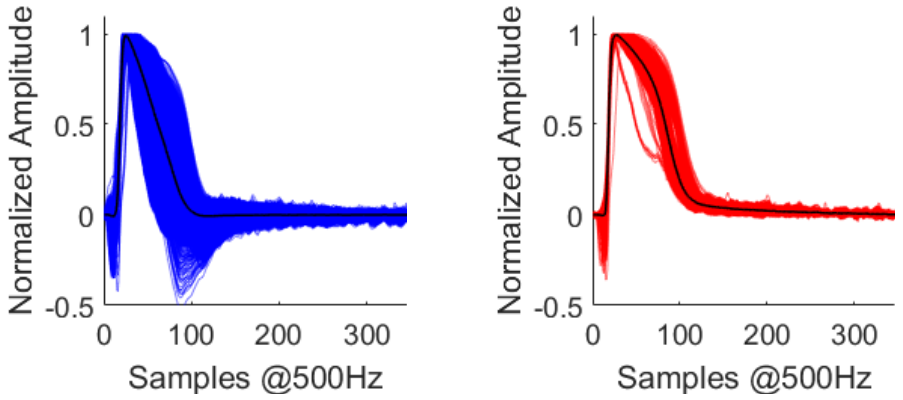
probability that each trial associates to the prediction. To do so, we computed the average output $\hat{y}$ over the trials per sample, and then we classified each sample as atrial-like if $\hat{y} < 0.5$, and ventricular-like if $\hat{y} \geq 0.5$. This can be seen as a soft-voting classification since it considers not only the class prediction of each trial, but also the probability associated to the prediction. This is what we refer to as *average prediction* of the network across trials[5] and it is used to analyze the performance of each one of the training approaches.

For the optical mapping dataset, Figure 2.11a and 2.11b show the atrial-like and ventricular-like population respectively[6], where the black curve in each plot represents the average AP per class. Figure 2.11c shows how the classes are distributed in the cell aggregates. As it can be seen, the supervised learning scheme generates spatially continuous classification regions even though it does not take into account spatial information.

For the single cell recording dataset, the output $\hat{y}$ at the last epoch was averaged over trials to obtain the *average prediction* of the network across trials. Histograms of these values per class (based on ground truth labels) are shown in Figure 2.12. Ideally, all the atrial-like samples would be below $\hat{y} = 0.5$ and all the ventricular-like samples would be above $\hat{y} = 0.5$. In this case most of the samples are just mapped close to 0, which leads to poor classification accuracy (32.69%). Nodal-like samples of the single cell recording dataset are

[5]Please note that the performance of the *average prediction* (which integrates information from different trials) need not correspond to the average performance over the trials.

[6]According to the *average prediction* of the network across trials.



(a) Atrial-like APs

(b) Ventricular-like APs

(c) Results in cell aggregates

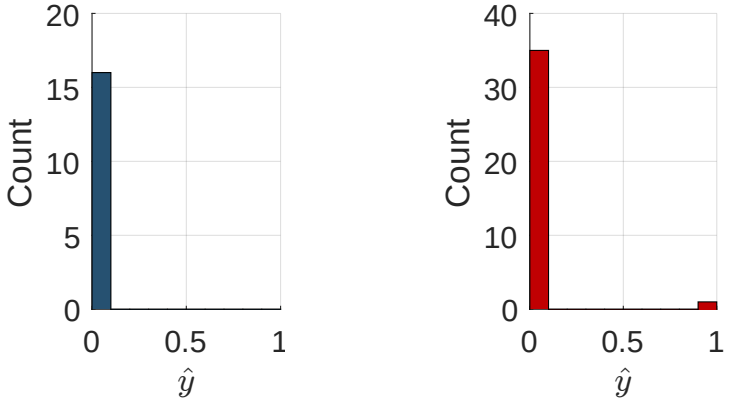
Figure 2.11: Results of the Sup-LSTM network in unlabeled hESC-CM dataset (DBI: 0.2834).

not included in the results since the network is trained only with atrial and ventricular samples [7].

The summary of the performance of the Sup-LSTM network, understood as the classification accuracy in the single cell recording dataset and DBI in the optical mapping dataset[8], is shown in Table 2.3 for each individual trial and for the *average prediction* of the network across trials. Please note that the performance of the *average prediction* of the network across trials need not correspond to the average performance over the trials.

[7]This is also the case for the results of all the approaches presented in this thesis.

[8]The DBI corresponds to the clustering quality index computed at the last epoch of training.



(a) Atrial-like hESC-CMs

(b) Ventricular-like hESC-CMs

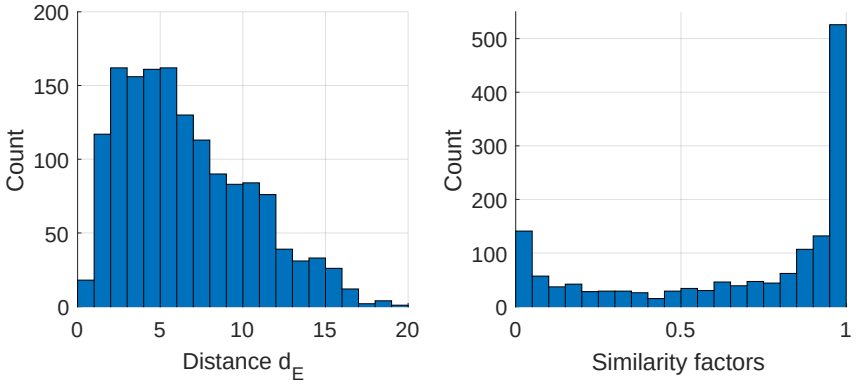
Figure 2.12: Results of the Sup-LSTM network in labeled hESC-CM dataset (Classification accuracy: 0.3269).

	Trial #1	**Trial #2**	**Trial #3**	**Trial #4**	**Trial #5**	**Average prediction**
Accuracy	0.3269	0.3269	0.3269	0.3269	0.3269	0.3269
DBI	0.2841	0.2830	0.2875	0.2826	0.2826	0.2834

Table 2.3: Summary performance of the Sup-LSTM network.

2.5.4.2 Semi-LSTM-E

In this case Euclidean distances between consecutive embryonic APs within the mini-batches are computed before training. Figure 2.13a shows the histogram of the 1500 computed distances, from which the parameter σ_s is obtained to compute the similarity factors as defined in (2.9). As their name suggests, similarity factors are used to indicate whether two samples are similar or not. Thus, they are more informative when they are close to 1 (similar samples) or 0 (dissimilar samples). As it can be seen in Figure 2.13b, the histogram of the resulting similarity factors $s_{(j,j')}$ exhibits peaks close to 1 and 0, which



(a) Histogram Euclidean distances training set

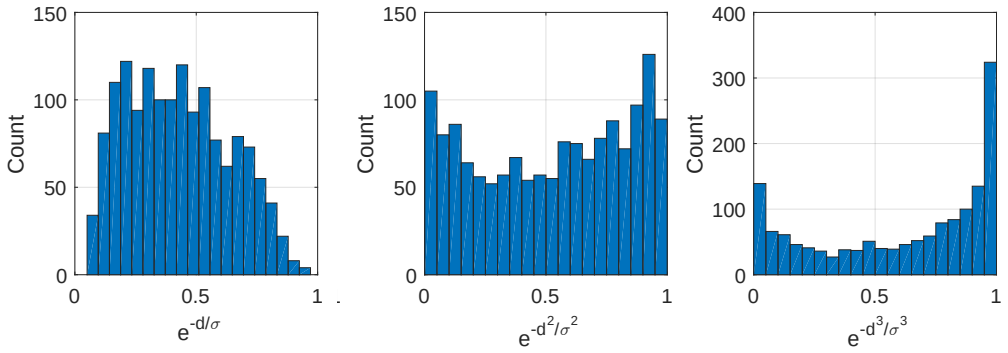
(b) Histogram similarity factors based on Euclidean distances

Figure 2.13: (a) Euclidean distances d_E, and (b) the corresponding similarity factors $s_{(j,j')}$ with $\sigma_s = 9.2597$.

supports the decision of using the fourth power of the distance when computing the similarity factors (as opposed to simply using a Gaussian function). Figure 2.14 shows how the histograms of the similarity factors would look like if the first, second or third power of the distances were used instead.

Figure 2.15 depicts the training results over the 200 epochs. The first column shows the loss function evaluated in the training set (top) and validation set (bottom). It can be seen that in both cases the loss function has a decreasing trend that becomes flat by the end of training.

The second column of Figure 2.15 shows that the supervised term as well as the unsupervised term converge to similar values across trials. It can be seen that the unsupervised term increases its value when there is a significant decrease in the supervised term, which is indicative of a compromise between both terms. However, as the optimization continues, the unsupervised term



(a) First power (b) Second power (Gaussian) (c) Third power

Figure 2.14: Histograms of the similarity factors if the (a) first, (b) second, or (c) third powers of the distances were used.

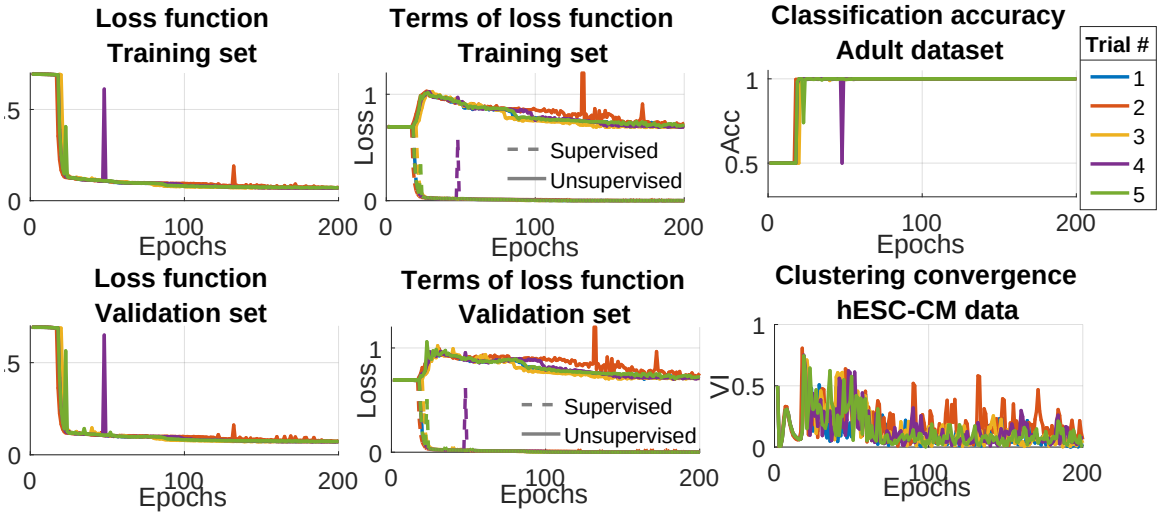


Figure 2.15: Training results of the Semi-LSTM-E network.

decreases its value while the supervised term stays close to zero.

The third column of Figure 2.15 shows that the network learns to correctly classify samples in the adult domain (classification accuracy close to 1). However in this case the evolution of the VI between consecutive epochs shows that

even when the loss function has converged, the clustering results are not absolutely stable. This means that different sets of network parameters lead to similar values in the loss function, but generate different clustering results. Actually, in Figure 2.16 the network parameters at the last epoch exhibit a large variability across trials.

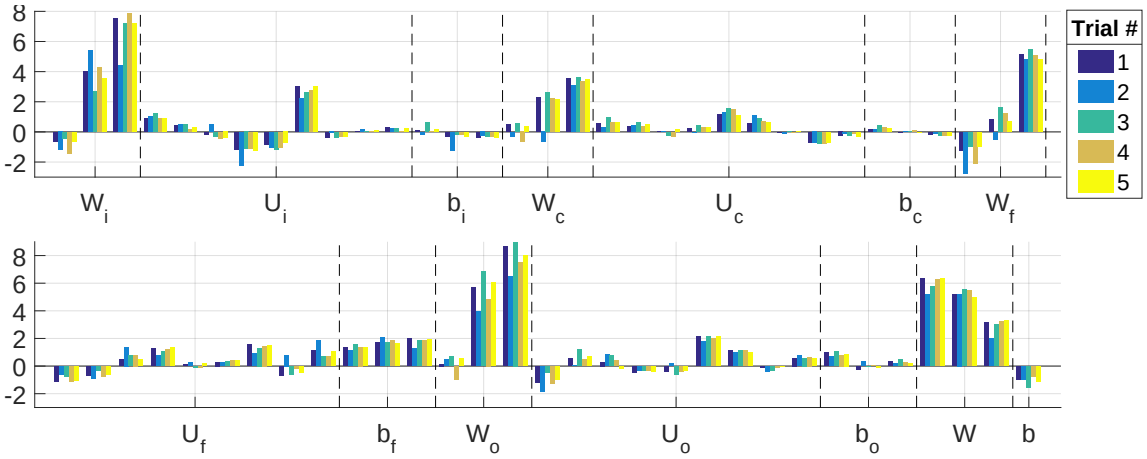
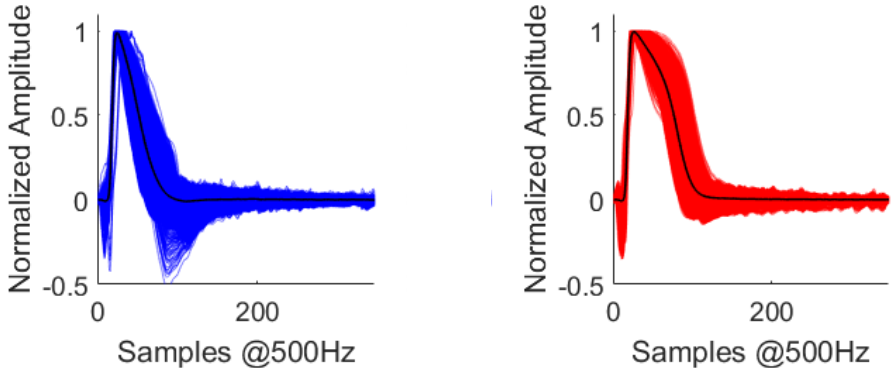


Figure 2.16: Weights of the Semi-LSTM-E network at the last epoch of training for five different trials.

The *average prediction* of the network across trials per sample was computed. Figure 2.17a shows the samples classified as atrial-like in the optical mapping dataset while Figure 2.17b shows the samples classified as ventricular-like. Figure 2.17c depicts the *average prediction* of the network across trials in the cell aggregates. As it can be seen, the Semi-LSTM-E network produces spatially smooth classification regions.

For the single cell recording dataset, the *average prediction* of the network across trials was computed. Figure 2.18 shows the histograms of these values



(a) Atrial-like APs

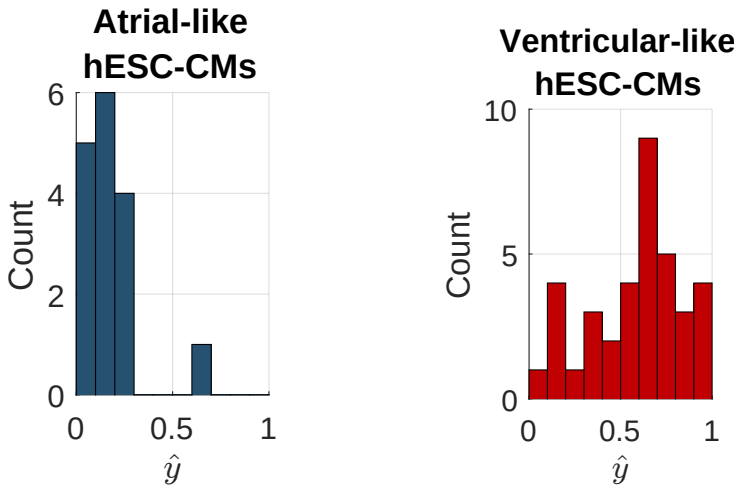
(b) Ventricular-like APs

(c) Results in cell aggregates

Figure 2.17: Results of the Semi-LSTM-E network in unlabeled hESC-CM dataset (DBI: 0.2458).

per class (based on ground truth labels). As it can be seen, most of the atrial-like samples are correctly classified ($\hat{y} < 0.5$), but many ventricular-like APs are misclassified (the histogram in Figure 2.23b spans across the whole range $\hat{y} \in [0, 1]$). In this case the classification accuracy reaches 76.92% for the *average prediction* across trials, since 1 atrial-like and 11 ventricular-like samples are misclassified.

Table 2.4 shows the summary of the performance per trial and for the *average prediction* of the network across trials. In this case the performance of the Semi-LSTM-E network in the *average prediction* is better than in most of the individual trials, which it is expected since the *average prediction* combines the



(a) Atrial-like hESC-CMs

(b) Ventricular-like hESC-CMs

Figure 2.18: Results of the Semi-LSTM-E network in labeled hESC-CM dataset (Classification accuracy: 0.7692).

information from different trials.

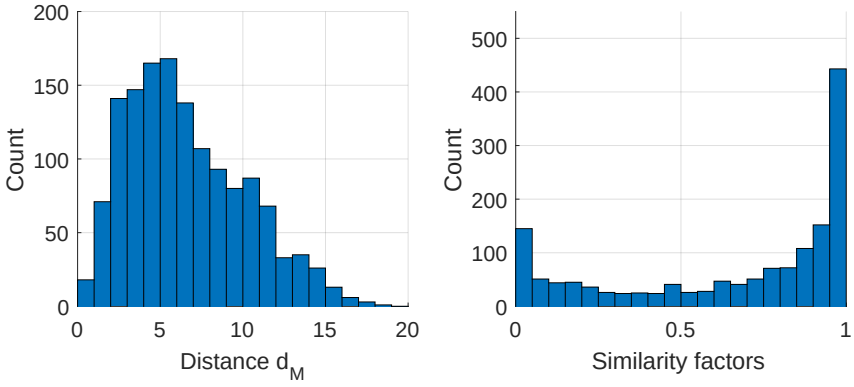
	Trial #1	**Trial #2**	**Trial #3**	**Trial #4**	**Trial #5**	**Average prediction**
Accuracy	0.75	0.75	0.7308	0.75	0.7692	0.7692
DBI	0.2466	0.2444	0.2469	0.2486	0.2476	0.2458

Table 2.4: Summary performance of the Semi-LSTM-E network.

2.5.4.3 Semi-LSTM-M

In this case, metamorphosis distances between consecutive embryonic APs within the mini-batches are computed before training. Figure 2.19a shows the histogram of the 1500 computed distances, from which the parameter σ_s is obtained to compute the similarity factors. As in the Semi-LSTM-E case, the

histogram of the resulting similarity factors $s_{(j,j')}$ shown in Figure 2.19b also exhibits peaks close to 1 and 0.



(a) Histogram metamorphosis distances training set

(b) Histogram similarity factors based on metamorphosis distances

Figure 2.19: (a) Metamorphosis distances d_M, and (b) the corresponding similarity factors $s_{(j,j')}$ with $\sigma_s = 8.9183$.

Figure 2.20 depicts the training results over the 200 epochs. In this case same observations as in the Semi-LSTM-E case hold: (i) the loss function shows a decreasing trend in all trials for training and validation sets; (ii) the unsupervised term exhibits an increase when the supervised term shows a abrupt decrease; (iii) in all trials the network learns to correctly classify adult samples; (iv) the clustering results do not converge; and (v) the variability between trials of the sets of parameters at last epoch shown in Figure 2.21 is large.

Figure 2.22a shows the samples classified as atrial-like in the optical mapping dataset, while Figure 2.22b shows the samples classified as ventricular-like[9]. Figure 2.22c depicts the *average prediction* of the network across trials

[9]According to the *average prediction* across trials.

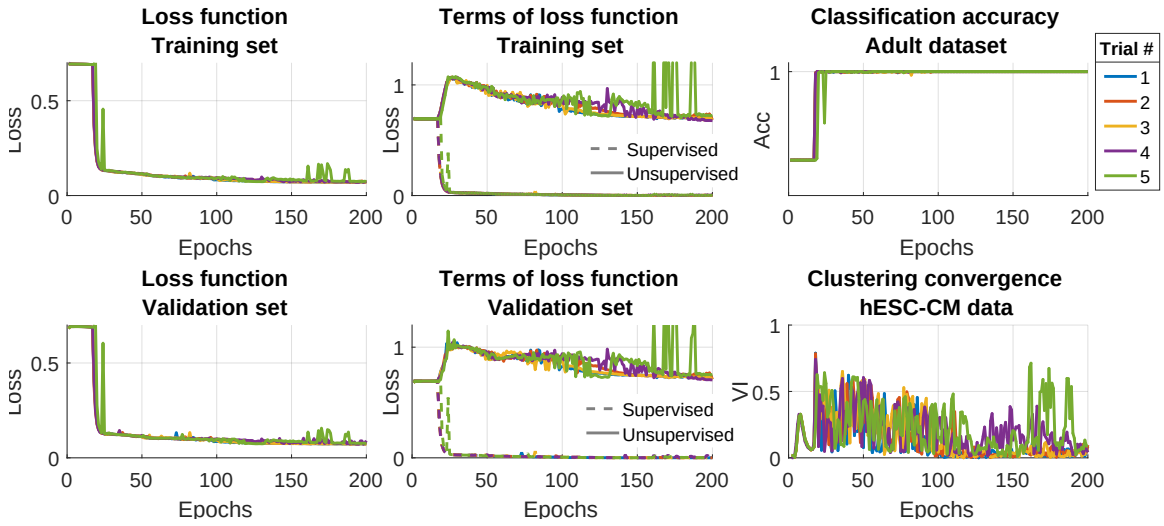


Figure 2.20: Training results of the Semi-LSTM-M network.

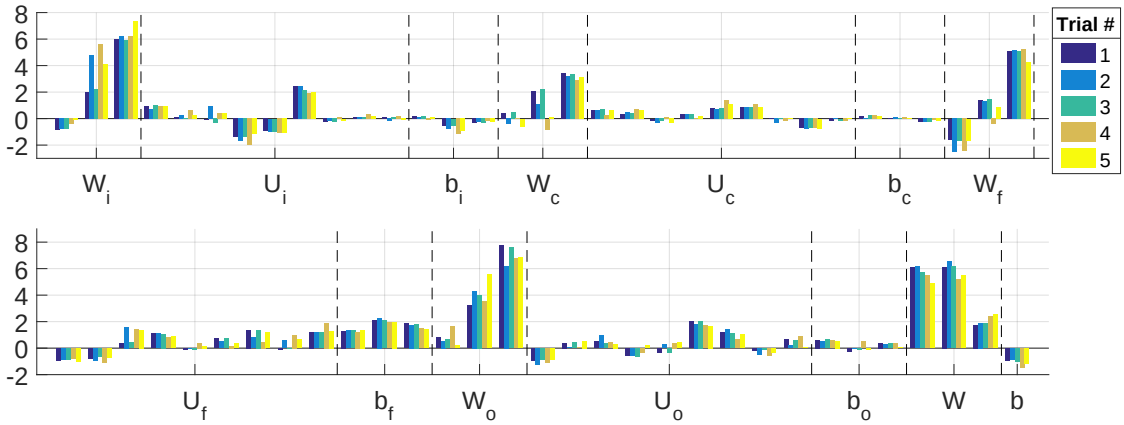
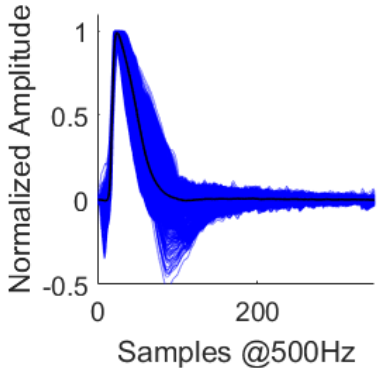


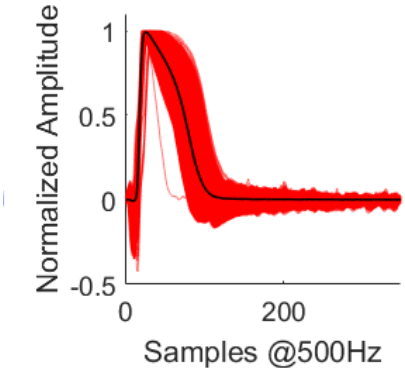
Figure 2.21: Weights of the Semi-LSTM-M network at the last epoch of training for five different trials.

in the cell aggregates.

For the single cell recording dataset, Figure 2.23 shows the histograms per class (based on ground truth labels) of the *average prediction* of the network across trials. In this case the classification accuracy is 76.92%, since 5 atrial-like and 7 ventricular-like samples are misclassified.



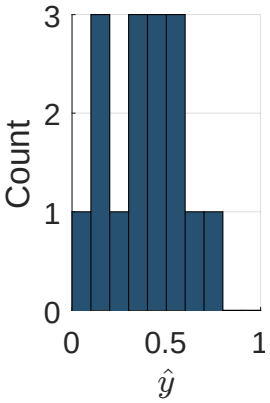
(a) Atrial-like APs



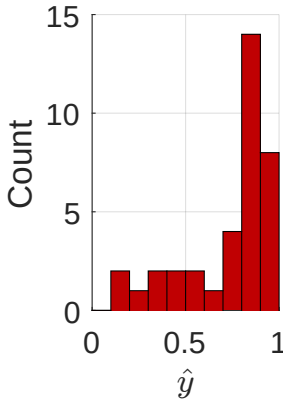
(b) Ventricular-like APs

(c) Results in cell aggregates

Figure 2.22: Results of the Semi-LSTM-M in unlabeled hESC-CM dataset (DBI: 0.2390).



(a) Atrial-like hESC-CMs



(b) Ventricular-like hESC-CMs

Figure 2.23: Results of the Semi-LSTM-M network in labeled hESC-CM dataset (Classification accuracy: 0.7692).

Table 2.5 shows the summary of the performance of the Semi-LSTM-M network per trial and for the *average prediction* of the network across trials. The

accuracy corresponds to the classification accuracy obtained in the single cell recording dataset (considering only atrial-like and ventricular-like samples), and the DBI is the clustering quality index obtained in the optical mapping dataset at the last epoch of training.

	Trial #1	**Trial #2**	**Trial #3**	**Trial #4**	**Trial #5**	**Average prediction**
Accuracy	0.7115	0.7308	0.7115	0.8077	0.7885	0.7692
DBI	0.2441	0.2414	0.2441	0.2379	0.2413	0.2390

Table 2.5: Summary performance of the Semi-LSTM-M network.

2.5.4.4 Effect of SGD with predefined random mini-batches

In mini-batch stochastic gradient descent method, a subset of samples (mini-batch) is used to estimate the gradient of the cost function. The classic approach consists of building a new random mini-batch from the training set at each iteration, however in our case that would require new metamorphosis computations every time, which is unfeasible. Alternatively, we propose to select the mini-batch from a predefined set of mini-batches that have been randomly built before training.

We conjecture that the proposed modification does not significantly affect the guarantees of the classic approach of mini-batch SGD. Therefore, we performed experiments to evaluate how the particular set of random mini-batches used for training affected the performance of the algorithm. Since computing

metamorphosis distances at every iteration of the optimization algorithm is not possible, the Semi-LSTM-E case was selected for this purpose. Four cases are studied:

(a) **No shuffle:** Mini-batches are randomly built at the beginning of training and are not updated between epochs. This corresponds to the proposed training scheme.

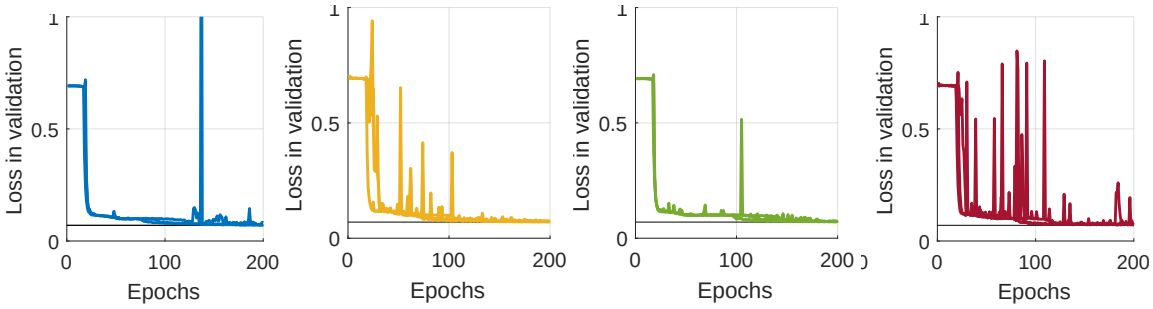
(b) **Shuffle only adult:** Mini-batches are built at the beginning of training, but the adult samples are shuffled across mini-batches after every epoch.

(c) **Shuffle only embryonic:** Mini-batches are built at the beginning of training, but the embryonic samples are shuffled across mini-batches after every epoch. Similarity factors must be re-computed every epoch.

(d) **Shuffle both:** Mini-batches are randomly built at the beginning of each epoch. Similarity factors must be re-computed every epoch. This corresponds to the classic mini-batch SGD approach.

The set of mini-batches used for validation is fixed across simulations. Three trials are simulated for each case, and the loss function evaluated in the validation set is shown in Figure 2.24, where the horizontal black line has been included as a reference. The validation loss indicates how well the network generalizes, and therefore the fact that it converges to a similar value in all the studied cases means that not significant bias is introduced with this finite set

of random mini-batches. Validation loss of "Shuffle both" case oscillates more than that of the other cases, which can be explained because it involves more randomness; therefore it requires more time to be learned by the network.

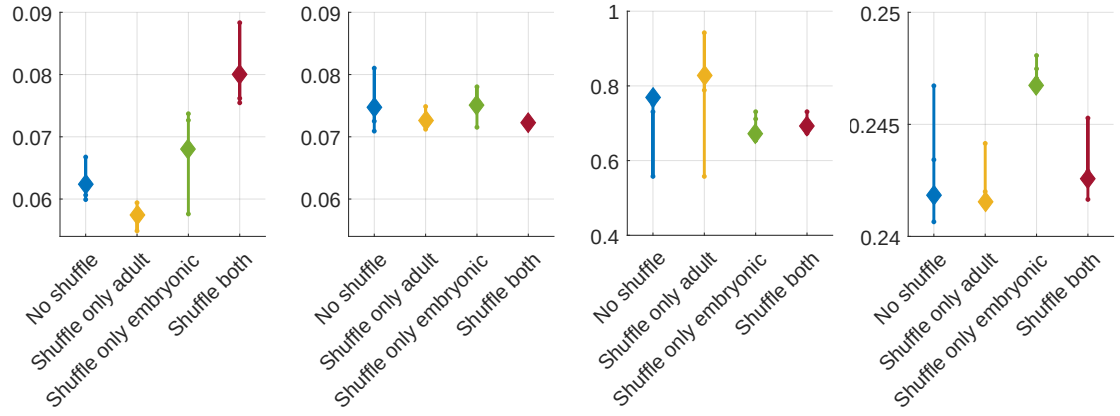


(a) No shuffle (b) Shuffle only adult (c) Shuffle only embryonic (d) Shuffle both

Figure 2.24: Validation loss different cases of mini-batch optimization.

The last value of the loss function in training set, the last value of the loss function in validation set, the classification accuracy in the single cell recording dataset and the DBI in the optical mapping dataset were computed and plotted in Figure 2.25 for each case. Although the training loss reaches different values on the different cases, there is not a significant difference in the average validation loss. However, the classic approach of mini-batch SGD ("Shuffle both") does generate slightly lower validation loss with significantly lower variance than the other cases.

Although the variance of the classification accuracy is also significantly smaller than the rest for the classic mini-batch SGD, the best classification accuracy is obtained by the "shuffle only adult" approach. Moreover, "shuffle



(a) Last value loss in training set

(b) Last value loss in validation set

(c) Accuracy in labeled hESC-CM dataset

(d) DBI in unlabeled hESC-CM dataset

Figure 2.25: Indicators for cases of mini-batch stochastic optimization. $\diamond$ is the average training and validation loss over trials in (a) and (b), respectively. $\diamond$ is the accuracy and DBI of the *average prediction* in (c) and (d), respectively. Vertical lines indicate the range covered by the trials.

only adult" case also obtains the best DBI in its *average prediction* of the optical mapping dataset, but it is within the range of variability of other cases.

However, the DBI observed in all the trials across cases is always better than the results observed for Sup-LSTM case and does not reach the average performance of Semi-LSTM-M case. Therefore, although the limited number of trials does not allow us to conclude on which approach works best for this given problem, we can at least conclude that the performance of the method is not significantly affected by this particular predefined set of mini-batches used for training.

2.5.5 Analysis

It is worth noting that in both supervised and semi-supervised approaches the proposed classifier learns to correctly classify samples in the adult domain. However, that problem can be solved by looking at simple AP features and it is not the task in which we are interested. The great performance achieved in the adult domain is not contradictory with a good performance in the embryonic domain. Given an AP x, we assume that the conditional probability of its class is independent of its domain. Thus, the goal of the classifier is to approximate a general function $\mathbb{P}\{y|\mathrm{x}\}$ that applies to both adult and embryonic samples.

Regarding classification of embryonic samples, Figure 2.26 and Table 2.6 compare our results to those of the method presented in [1] (1NN classifier with $N_a = 20$ synthetic adult AP templates). In all the proposed approaches the classifier generates spatially smooth classification regions and suggests heterogeneity in most of the cell aggregates, which coincides with previous findings [1, 27].

Method	1NN	1NN	1NN	Sup-LSTM	Semi-LSTM	Semi-LSTM
Templates	20 [1]	20 [1]	300 SMRS	300 SMRS	300 SMRS	300 SMRS
Metric	M	E	E		E	M
Accuracy	**0.9615**	0.8654	0.8077	0.3269	0.7692	0.7692
DBI	**0.2297**	0.2558	0.2566	0.2834	0.2458	0.2390
Accuracy*	N/A	0.6488	0.6290	0.4723	0.8788	**0.9473**

Table 2.6: Comparing the results of the proposed semi-supervised method with the results presented in [1]. Accuracy is computed in single cell recording hESC-CM dataset. DBI is computed in optical mapping dataset. Accuracy* is computed in optical mapping dataset assuming 1NN classification with metamorphosis distance as ground truth (E: Euclidean, M: Metamorphosis).

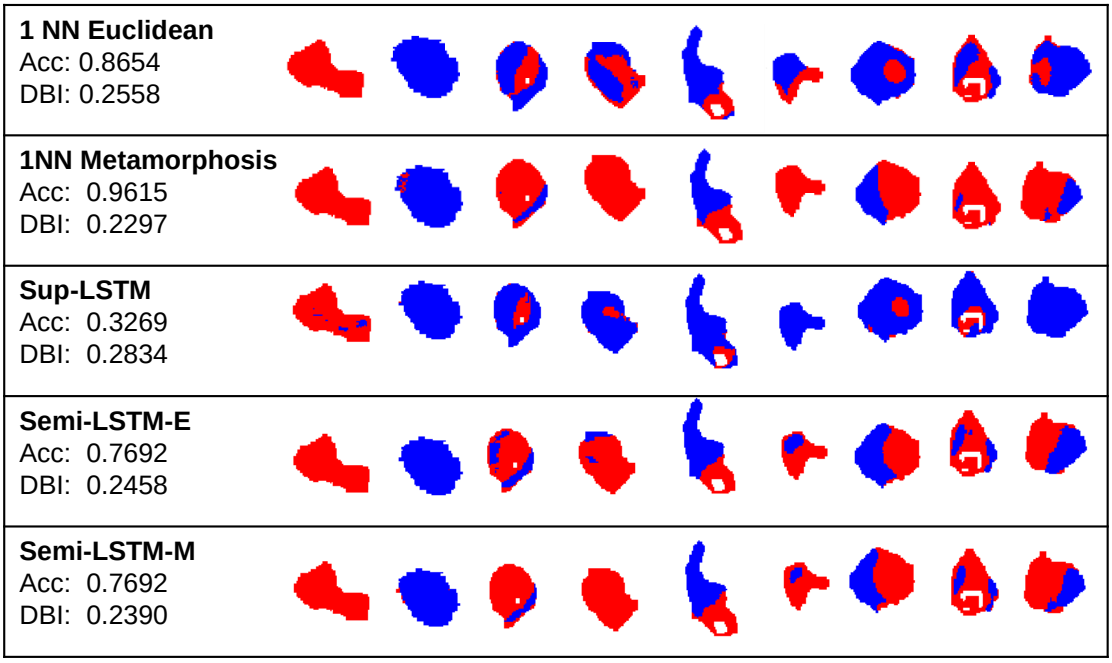


Figure 2.26: Comparison of the results of semi-supervised approach with the results presented in [1]. Acc corresponds to the classification accuracy in the single cell recording dataset, and DBI is the clustering quality index in the optical mapping dataset.

As it can be observed in Figure 2.26, the semi-supervised approach (with Euclidean and with metamorphosis distances) generates classification results that are visually similar to the state of the art (1NN method with metamorphosis distances), while the classification results of the Sup-LSTM network are remarkably different. Besides, the proposed semi-supervised approach significantly outperforms the supervised approach in terms of clustering quality and classification accuracy. This emphasizes that adult and embryonic APs intrinsically belong to different domains, and therefore classifying embryonic APs with a network trained only with adult APs is not adequate. Moreover,

it validates the proposed contrastive unsupervised loss as a suitable way of incorporating information from unlabeled embryonic samples in training.

Although the similarity factors do not look significantly different when they are computed based on Euclidean or metamorphosis distances (see Figures 2.13 and 2.19), the use of metamorphosis distances does generate better clustering quality results not only in the *average prediction* but also in individual trials (compare Tables 2.4 and 2.5). Therefore, in this framework there exists an important trade-off between computational cost of similarity factors[10] and clustering quality. However, this trade-off does not seem to be relevant in their transferability to a new dataset, since both semi-supervised approaches obtain the same classification accuracy in the single cell recording dataset (76.92%).

It can be seen in Table 2.6 that supervised learning shows significantly higher DBI than the rest, which is not surprising since it does not consider hESC-CM data during training. On the other hand, the semi-supervised learning scheme outperforms the 1NN scheme when Euclidean distances are used (DBI 0.2458 vs 0.2558). 1NN with Euclidean distances was replicated with the same 300 adult AP templates used to train the network (see Table 2.6)[11], confirming that the improvement in clustering quality observed in the semi-supervised scheme is not attributable to the number of templates used, but to

[10]Cost of computing the Euclidean distances versus cost of computing the metamorphosis distances.

[11]It is unclear how the 20 templates were selected by the authors in [1]. Therefore, the decrease in performance observed when 300 templates are used could be attributable to the handpicked selection of the 20 templates.

the method itself: The Euclidean metric is a good approximation of metamorphosis for small distances, so it performs better when distances within hESC-CM domain are computed (proposed semi-supervised framework) than when distances between hESC-CM and adult CM domains are computed (1NN).

1NN metamorphosis results presented in [1] show the best clustering quality (DBI 0.2297), followed by the Semi-LSTM-M (DBI 0.2390). Since no ground truth labels are available for this dataset, we compute the classification accuracy (Accuracy*) assuming 1NN metamorphosis as the ground truth in order to quantify how similar our predicted labels are to the ones provided by the state-of-the-art method.

The classification accuracy assuming 1NN metamorphosis as the ground truth[12] was computed and plotted versus the DBI in Fig. 2.27. In the optical mapping dataset, the use of metamorphosis distance in semi-supervised learning not only produces lower DBI but also consistently generates better classification accuracy than when the Euclidean distance is used (small dots in Fig. 2.27 represent single trials results and squares represent the *average prediction* per case). An improvement of 24.98% in the accuracy* is observed between 1NN (62.90%) and the semi-supervised learning scheme when 300 templates and only Euclidean distances are used, achieving 87.88% accuracy* without any metamorphosis distance computation.

[12]This assumption is based on the fact that it is the method that generates better clustering quality (lower DBI).

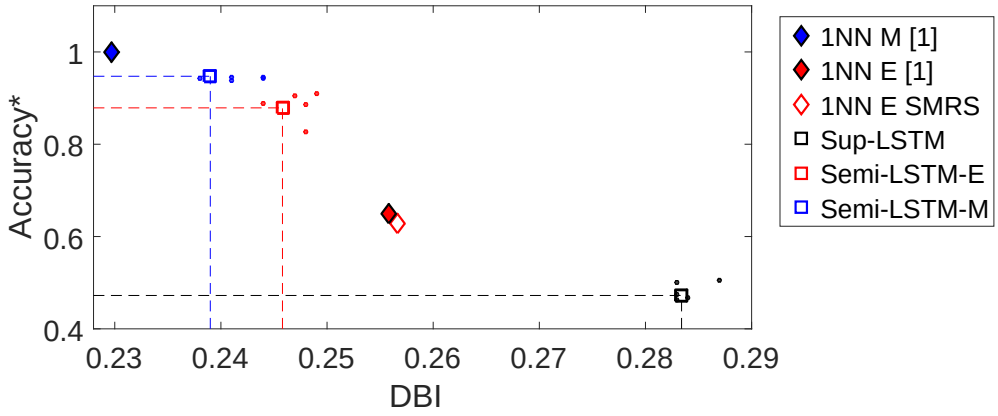


Figure 2.27: Accuracy* vs DBI. 1NN M as ground truth (E: Euclidean, M: Metamorphosis).

Therefore, the proposed method not only successfully integrates labeled data from a different domain to solve the task, but also proves to be a powerful framework to improve the performance of Euclidean-based methods in the classification of hESC-CM APs. Moreover, it reaches 94.73% of agreement with the state of the art, trading off accuracy with computational complexity: whereas the classification of a new sample in the state-of-the-art method requires the solution of 20 computationally intensive optimization problems (6.74 sec/sample in 2 8-core computer nodes with 8 2.3GHz CPUs per node [1]), in the proposed method it just needs to be processed by a small RNN with fixed weights (less than 6 sec for the whole 6940 AP dataset in one 2.2 GHz CPU with 2 cores, 4 threads)[13].

[13]The reported time corresponds to testing a dataset once the network has been trained. Training the network for 200 epochs takes approx. 30 minutes.

2.6 Chapter summary

In this chapter we have presented a new approach to classification of hESC-CMs, in which an RNN with LSTM units is trained using a semi-supervised loss function. Experimental results show that the proposed semi-supervised approach significantly outperforms a supervised approach when labeled data is only available for adult CMs. Moreover, it exhibits important computational advantages and achieves 94.73% of agreement with respect to the state of the art.

Chapter 3

A domain adaptation approach to classification of APs

One of the main assumptions of many machine learning algorithms is the fact that training and test data are sampled from the same probability distribution. Unfortunately, that is not always the case. In applications where ground truth labeling is an expensive and time-consuming procedure, one might want to use any available labeled data for training, even when it belongs to a different domain. In sentiment classification of customer feedback, for example, millions of public reviews of different products and services are easily available, but labeling enough data to cover every possible domain, including books, restaurants, computers, movies, etc., would require a lot of effort. On the other hand, it is reasonable to argue that labels on book reviews are informative to

classify reviews of movies, for example. But how do we use labels from one domain to make predictions for other domains?

Domain adaptation precisely addresses the problem of optimizing the performance in one domain (called target domain), given training data in a different domain (called source domain). While domain adaptation is conceptually different from multi-task learning, where performance is optimized for multiple tasks in multiple domains simultaneously, in both cases information is transferred between different domains, so they both fall under a broader concept called transfer learning [59].

Unsupervised domain adaptation assumes labeled data available only in the source domain to optimize performance in the target domain, as opposed to supervised domain adaptation, where labels from both domains are used. In this chapter we present an unsupervised domain adaptation approach[1] to train a recurrent neural network for classifying hESC-CM APs. In our case the target domain corresponds to hESC-CM APs, whereas the source domain is characterized by adult CM data for which ground truth labels are available.

[1]Although unsupervised domain adaptation is technically a semi-supervised method because it uses labeled and unlabeled data, we reserve the term "semi-supervised approach" for the learning method presented in Chapter 2.

3.1 Problem formulation

Let $\Omega_e = \{\mathbf{x}_j^e\}_{j=1}^{N_e}$ be an *unlabeled* dataset from the target domain, where the sequence $\mathbf{x}_j^e = \{x_j^e(k) \in \mathbb{R}\}_{k=1}^{K}$ represents the jth embryonic AP and K is the total number of samples in one cycle length. Let $\Omega_a = \{(\mathbf{x}_i^a, y_i^a)\}_{i=1}^{N_a}$ be a *labeled* adult dataset from the source domain, where $\mathbf{x}_i^a = \{x_i^a(k) \in \mathbb{R}\}_{k=1}^{K}$ is the ith adult AP and $y_i^a \in \{0, 1\}$ is its ground truth label ($y_i^a = 0$ denotes atrial and $y_i^a = 1$ denotes ventricular). We consider the problem of assigning a label $\hat{y}_j^e$ to each $\mathbf{x}_j^e \in \Omega_e$, where $\hat{y}_j^e = 0$ denotes atrial-like and $\hat{y}_j^e = 1$ denotes ventricular-like.

Let $\mathbb{P}_e\{\mathbf{x}\}$ be the probability density function of AP $\mathbf{x}$ in the embryonic domain and $\mathbb{P}_a\{\mathbf{x}\}$ be the probability density function of AP $\mathbf{x}$ in the adult domain. The underlying probability density functions $\mathbb{P}_e\{\mathbf{x}\}$ and $\mathbb{P}_a\{\mathbf{x}\}$ are assumed to be unknown, but we assume that the probability distributions of the embryonic and adult domains are different, i.e., $\mathbb{P}_e \neq \mathbb{P}_a$. Besides, we assume covariate shift which means that for a given AP $\mathbf{x}$, the conditional probability of its class is independent of its domain, i.e., $\mathbb{P}_e\{y|\mathbf{x}\} = \mathbb{P}_a\{y|\mathbf{x}\} = \mathbb{P}\{y|\mathbf{x}\}$.

According to [33], domain adaptation approaches to this problem can be classified as: (i) Instance weighting approaches, in which source examples are weighted in training in order to resemble the target distribution; (ii) Self-labeling approaches, in which an initial guess of the labels in the target domain

is generated based on labeled source data, and iteratively modified according to relationships within target domain; (iii) clustering-based methods, in which labels from the source domain are transferred to target domain based on similarities between inter-domain samples; or (iv) feature representation methods, in which a subset of the original feature space or a new feature representation is used to capture shared characteristics of both domains.

Instance weighting approaches require shared support between both distributions, i.e., $\forall \mathbf{x}, \mathbb{P}_a\{\mathbf{x}\} = 0$ iff $\mathbb{P}_e\{\mathbf{x}\} = 0$. In our case there are embryonic APs that are never observed in adult data, therefore the shared support assumption does not hold and instance weighting approaches are discarded. On the other hand, self-labeling approaches as well as clustering-based methods often rely on computing similarities between samples, which in the case of APs can be computationally expensive. Thus, we follow a feature representation approach to unsupervised domain adaptation, in which probability distribution functions of both domains are forced to be similar in a different feature space $\varphi(\mathbf{x})$.

3.2 Maximum Mean Discrepancy

Maximum mean discrepancy (MMD) is a statistic presented by Gretton et al. in 2007 as an approach to design statistical tests to determine if two samples are drawn from different distributions [46]. It corresponds to the distance

between the mean of the two samples mapped into a reproducing kernel Hilbert space (RKHS). An estimation of the MMD between two datasets $\Omega_a = \{\mathbf{x}_i^a\}_{i=1}^{N_a}$ and $\Omega_e = \{\mathbf{x}_j^e\}_{j=1}^{N_e}$ is given by

$$\widehat{MMD}\left(\Omega_a, \Omega_e\right) = \left\| \frac{1}{N_a} \sum_{i=1}^{N_a} \psi(\mathbf{x}_i^a) - \frac{1}{N_e} \sum_{j=1}^{N_e} \psi(\mathbf{x}_j^e) \right\|_{\mathcal{H}}, \tag{3.1}$$

where $\psi(\mathbf{x})$ is the embedding of $\mathbf{x}$ into the RKHS. After defining the positive semidefinite kernel as $\mathcal{K}(x, y) = \psi(x)^\top \psi(y)$, the MMD estimate shown in (3.1) can be rewritten as

$$\widehat{MMD}^2\left(\Omega_a, \Omega_e\right) = \left[\frac{1}{N_a^2} \sum_{i=1}^{N_a} \sum_{i'=1}^{N_a} \mathcal{K}\left(\mathbf{x}_i^a, \mathbf{x}_{i'}^a\right) - \frac{2}{N_a N_e} \sum_{i=1}^{N_a} \sum_{j=1}^{N_e} \mathcal{K}\left(\mathbf{x}_i^a, \mathbf{x}_j^e\right) + \frac{1}{N_e^2} \sum_{j=1}^{N_e} \sum_{j'=1}^{N_e} \mathcal{K}\left(\mathbf{x}_j^e, \mathbf{x}_{j'}^e\right) \right] \tag{3.2}$$

A commonly used kernel is the Gaussian kernel given by

$$\mathcal{K}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \exp\left(\frac{-\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{2\sigma_k^2} \right), \tag{3.3}$$

where σ_k is a design parameter.

When the probability densities $\mathbb{P}_e$ and $\mathbb{P}_a$ are unknown, the MMD estimate in (3.2) allows us to estimate how different they are based on their samples. The MMD estimator has been successfully applied to learn appropriate kernels for cross-domain SVM-based classification, regression and video concept detec-

tion, among others [60, 61]. The estimator has also been recently applied with fixed kernels as a metric to learn the parameters of generative networks [62], and the parameters of feature extraction layers for multi-task learning in multiple domains [63], which is closely related to our task.

3.3 A domain adapted LSTM approach to classification of APs

As discussed in Chapter 2, LSTMs are a promising candidate for classification of time-series in general, and consequently for classification of hESC-CM APs in particular. However, one disadvantage of LSTMs is that they do not explicitly model the domain shift between embryonic and adult APs. In this section we propose a domain adapted LSTM approach to classification of APs. In the proposed approach one of the hidden layers of the LSTM is used to define a shared feature between source and target domains (see Section 3.3.1) and the MMD loss between the distribution of these features is used to reduce the domain shift (see Section 3.3.2).

3.3.1 An RNN with LSTM units as feature extractor and classifier

The architecture of the proposed classifier is depicted in Figure 3.1 and consists of one input layer, one hidden LSTM layer of dimension $p = 3$, and a single sigmoid unit as the output layer. As in previous chapter, let $\mathcal{W} = \{W_i, W_f, W_o, W_c, U_i, U_f, U_o, U_c, b_i, b_f, b_o, b_c, W, b\}$ be the set of parameters of the proposed classifier. In this approach the LSTM layer is explicitly considered as a feature extractor, such that $\mathbf{x} \mapsto \varphi_{\mathcal{W}}(\mathbf{x}) = h(\mathbf{x}, K) \in \mathbb{R}^3$, where the dependency of the output cell $h(\cdot)$ on the input sequence $\mathbf{x}$ is made explicit, and K corresponds to its last time step[2].

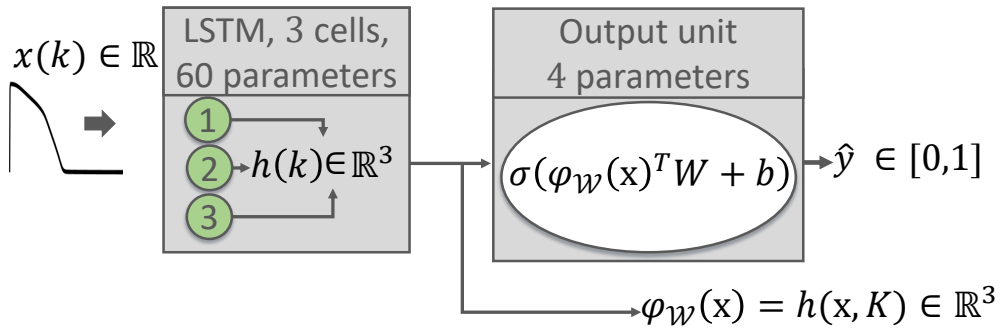


Figure 3.1: Proposed domain adapted classification approach.

The classification itself is carried out by the sigmoid unit that operates only on the last value of the hidden layer output, i.e., the feature vector $\varphi_{\mathcal{W}}(\mathbf{x})$. The predicted label then is given by $\hat{y} = \sigma\left(\varphi_{\mathcal{W}}(\mathbf{x})^T W + b\right)$, where $W \in \mathbb{R}^3$ and $b \in \mathbb{R}$

[2]Subindex $\mathcal{W}$ on $\varphi_{\mathcal{W}}(\mathbf{x})$ is used to emphasize that the feature vector depends on some of the parameters of the network.

are parameters of the output unit, and $\sigma(z) = \dfrac{1}{1+e^{-z}}$ corresponds to the pointwise sigmoid function. For a given set of parameters $\mathcal{W}$, we will represent the classifier as the function $f_{\mathcal{W}}(\mathbf{x}) = \hat{y}$ that maps an action potential x to a predicted label $\hat{y}$.

Similar to the architecture presented in Chapter 2, in this case the number of parameters is also small (64 parameter in total) to avoid overfitting. However, the architecture design can be easily extended to address cases more complex than binary classification.

3.3.2 Proposed domain adapted loss function

The basic idea behind the loss function we use is to enforce similarity between the probability density functions of the source and target domains in feature space, i.e. $\mathbb{P}_a\{\varphi_{\mathcal{W}}(\mathbf{x})\} \approx \mathbb{P}_e\{\varphi_{\mathcal{W}}(\mathbf{x})\}$, while training a classifier with source domain data. Thus, the network learns to classify the samples in a feature space in which embryonic and adult data "are similar".

The proposed objective function is

$$\begin{aligned}&\frac{1}{N_a}\sum_{i=1}^{N_a}\ell\left(y_i^a, f_{\mathcal{W}}(\mathbf{x}_i^a)\right) + \gamma\left[\frac{1}{N_a^2}\sum_{i=1}^{N_a}\sum_{i'=1}^{N_a}\mathcal{K}\left(\varphi_{\mathcal{W}}(\mathbf{x}_i^a), \varphi_{\mathcal{W}}(\mathbf{x}_{i'}^a)\right) + \right.\\ &\left. -\frac{2}{N_a N_e}\sum_{i=1}^{N_a}\sum_{j=1}^{N_e}\mathcal{K}\left(\varphi_{\mathcal{W}}(\mathbf{x}_i^a), \varphi_{\mathcal{W}}(\mathbf{x}_j^e)\right) + \frac{1}{N_e^2}\sum_{j=1}^{N_e}\sum_{j'=1}^{N_e}\mathcal{K}\left(\varphi_{\mathcal{W}}(\mathbf{x}_j^e), \varphi_{\mathcal{W}}(\mathbf{x}_{j'}^e)\right)\right],\end{aligned} \tag{3.4}$$

where $\gamma \geq 0$ is a parameter that modulates the importance given to domain

adaptation in the optimization. The first term of (3.4) corresponds to the supervised loss presented in the previous chapter, where $\ell(y, \hat{y})$ is the binary crossentropy loss in (2.6), which measures how close the predicted label $\hat{y}$ is to the true label y. The domain adaptation part of the objective function corresponds to the empirical estimator of MMD in the feature space, where $\mathcal{K}(\cdot, \cdot)$ represents the Gaussian kernel presented in (3.3).

In this way, when the proposed loss function for domain adaptation is minimized, not only is the distance between predicted labels and ground truth labels minimized, but so is the distance between the probability density functions of the source and target domains in feature space.

We integrate the semi-supervised approach presented in the previous chapter with the proposed domain adaptation approach via the following general loss function

$$\begin{aligned}
&\frac{1-\lambda}{N_a}\left(\sum_{i=1}^{N_a}\ell\left(y_i^a, f_{\mathcal{W}}(\mathbf{x}_i^a)\right)\right) + \frac{\lambda}{N_e\left(N_e-1\right)}\left(\sum_{j=1}^{N_e}\sum_{j'\neq j}\ell_u\left(f_{\mathcal{W}}(\mathbf{x}_j^e), f_{\mathcal{W}}(\mathbf{x}_{j'}^e)\right)\right) \\
&+\gamma\left[\frac{1}{N_a^2}\sum_{i=1}^{N_a}\sum_{i'=1}^{N_a}\mathcal{K}\left(\varphi_{\mathcal{W}}(\mathbf{x}_i^a), \varphi_{\mathcal{W}}(\mathbf{x}_{i'}^a)\right) - \frac{2}{N_a N_e}\sum_{i=1}^{N_a}\sum_{j=1}^{N_e}\mathcal{K}\left(\varphi_{\mathcal{W}}(\mathbf{x}_i^a), \varphi_{\mathcal{W}}(\mathbf{x}_j^e)\right)\right. \\
&\left.+\frac{1}{N_e^2}\sum_{j=1}^{N_e}\sum_{j'=1}^{N_e}\mathcal{K}\left(\varphi_{\mathcal{W}}(\mathbf{x}_j^e), \varphi_{\mathcal{W}}(\mathbf{x}_{j'}^e)\right)\right],
\end{aligned} \tag{3.5}$$

where λ is the balancing parameter between supervised and unsupervised terms, and γ modulates the importance given to the domain adaptation term.

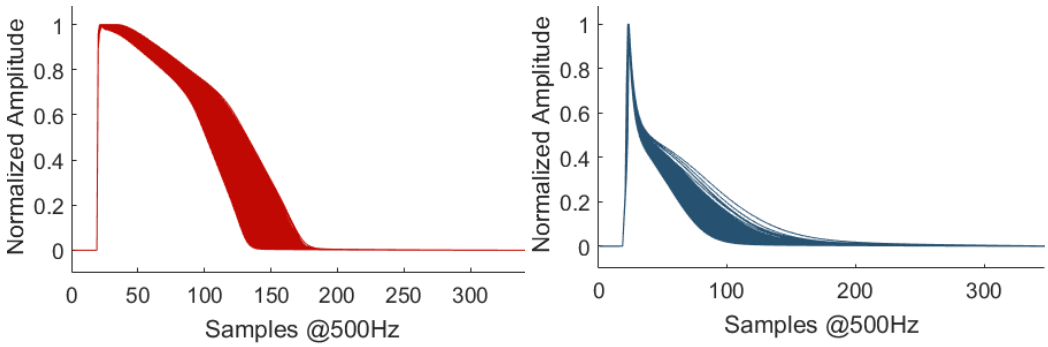
3.4 Metrics to evaluate classification and clustering

The set of metrics used to evaluate the performance of this approach corresponds to: (i) classification accuracy when labels are available for test data; (ii) Davies-Bouldin index (DBI) as a measure of clustering quality for unlabeled datasets; and (iii) variation of information (VI) to evaluate the convergence of the clusterings provided by the network during training. They are presented and described in detail in Section 2.4.

3.5 Experiments

3.5.1 Adult CM data

The population of 1000 ventricular and 1000 atrial adult APs generated by the procedure described in Section 2.5.1 is used for training and validation. However in this case, since adult samples are used to estimated the MMD statistic, we decided to increase the number of representatives selected by SMRS from 150 to 800 APs per class. Figure 3.2 shows the normalized APs selected for both classes.



(a) 800 normalized ventricular APs (b) 800 normalized atrial APs

Figure 3.2: Adult CM data: (a) A subset of 800 normalized synthetic ventricular examples generated by the ORd model [20] and selected by the SMRS method [54], and (b) a subset of 800 normalized synthetic atrial examples generated by the Nygren model [19] and selected by the SMRS method [54].

3.5.2 hESC-CM data

The two datasets available and described in Section 2.5.2 are used:

- **Labeled single cell recording dataset:** 16 atrial-like, 24 nodal-like and 36 ventricular-like APs (see Figure 2.6) used for testing.
- **Unlabeled optical mapping dataset:** 6940 APs coming from 9 cell aggregates. 1600 used for training and validation, and 5340 used for testing (see Figure 2.7).

3.5.3 Implementation details

The classifier architecture was implemented in Keras [57] with TensorFlow backend and trained using the RMSProp optimizer (with initial learning rate

$\epsilon = 0.003$). The network was initialized according to the default methods: states of the LSTM layer were initialized orthogonally, the forget bias was set as $b_f = 1$ [48], and the rest of the weights were initialized by the Glorot and Bengio method [58].

The 800 representatives of ventricular APs and the 800 representatives of atrial APs formed a dataset of $N_a = 1600$ adult samples. On the other hand, the training and validation subset of the optically mapped hESC-CMs formed a dataset of $N_e = 1600$ embryonic samples. A set of 100 random mini-batches, each one formed of 32 samples ($n_a = 16$ adults and $n_e = 16$ embryonic), was built at the beginning of training [3]. Therefore, the gradient of the supervised part of the loss function is estimated by the average of the $n_a = 16$ observations at each iteration, and the MMD is estimated in every iteration using 16 examples from each domain.

90 mini-batches were used for training, and 10 for validation. Figure 3.3 illustrates the training scheme described above. One epoch is considered to be a complete pass of the training dataset, which in our case corresponds to 90 iterations of the optimization algorithm.

Three cases are studied: domain adaptation on supervised learning (DA-

[3]For domain adaptation itself no metamorphosis computation is needed; therefore we could build random mini-batches during training. However, we decided to use a finite set of mini-batches because (i) we want to make sure that the differences of performance with respect to the semi-supervised approach are not attributable to the way in which the mini-batches are generated; and (ii) we want to integrate both approaches and in that case the generation of mini-batches has to be limited.

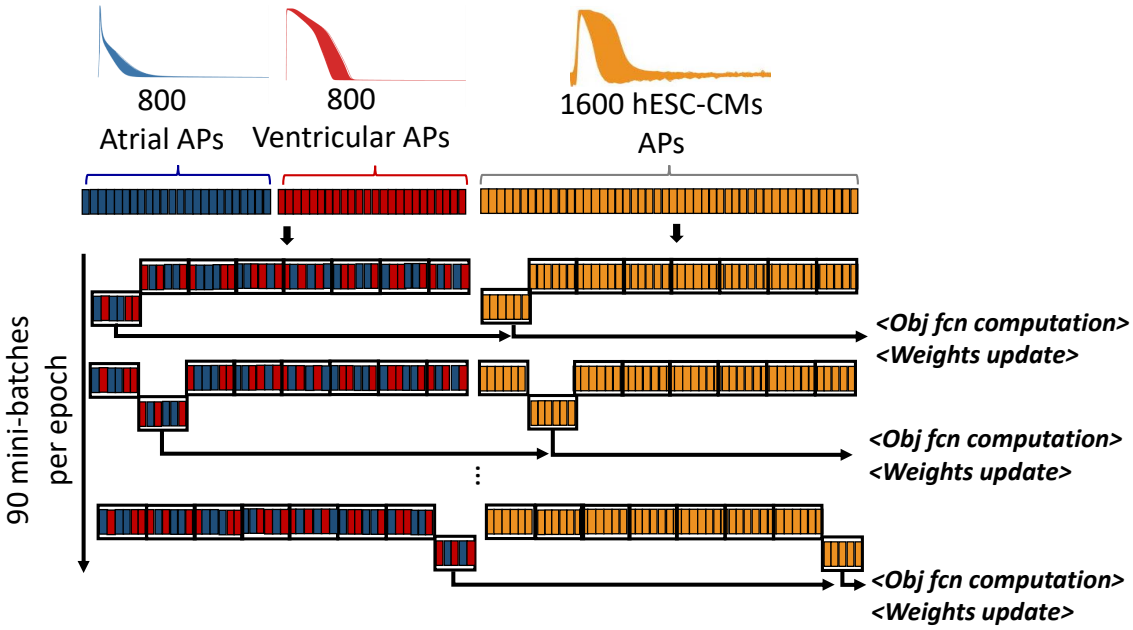


Figure 3.3: Training scheme for the domain adaptation approach.

Sup-LSTM), domain adaptation on semi-supervised learning with Euclidean distances (DA-Semi-LSTM-E), and domain adaptation on semi-supervised learning with metamorphosis distances (DA-Semi-LSTM-M). Different values of γ and different training conditions were explored, and the ones that generated best results per case are presented:

- **DA-Sup-LSTM.** The network was trained in a two step process: first 100 epochs with $\gamma = 0$ and $\lambda = 0$ as a Sup-LSTM initialization stage, and then 100 additional epochs with $\lambda = 0$ and $\gamma = 10$.

- **DA-Semi-LSTM-E.** The network was trained for 200 epochs with $\gamma = 1$ and $\lambda = 0.1$.

- **DA-Semi-LSTM-M.** The network was trained in a two step process: first

100 epochs with $\gamma = 0$ and $\lambda = 0$ as a Sup-LSTM initialization stage, and then 100 additional epochs with $\lambda = 0.1$ and $\gamma = 5$. The metamorphosis parameter was set as $\sigma_M = 0.3$.

In each case the network was trained 5 times with the same initialization for the weights. Each one of these 5 runs of the optimization algorithm is referred to as a "trial". The variability observed across trials for a given method (DA-Sup-LSTM, DA-Semi-LSTM-E or DA-Semi-LSTM-M) is attributable to the stochastic implementation of the optimization algorithm by Keras with TensorFlow backend. The kernel parameter was set as $\sigma_k = 0.3$.

3.5.4 Results

In this section the results of the DA-Sup-LSTM, DA-Semi-LSTM-E, and DA-Semi-LSTM-M networks are presented individually. A comparative analysis of their performances with respect to the performance of the networks introduced in the previous chapter (Sup-LSTM, Semi-LSTM-E and Semi-LSTM-M) and with respect to the state of the art (1NN classification with metamorphosis distances) is presented in Section 3.5.5.

3.5.4.1 DA-Sup-LSTM

Figure 3.4 presents the training results for the domain adaptation approach on supervised learning. As it can be seen, abrupt changes in the loss function are observed at epoch 100, which is expected since the training scheme changes from the initialization stage (Sup-LSTM) to domain adaptation itself (DA-Sup-LSTM). It is interesting to note in the second column of Figure 3.4 that although the domain adaptation term of the loss function decreases compared to the initialization stage in training set (top), it increases in the validation set (bottom), which can be indicative of overfitting.

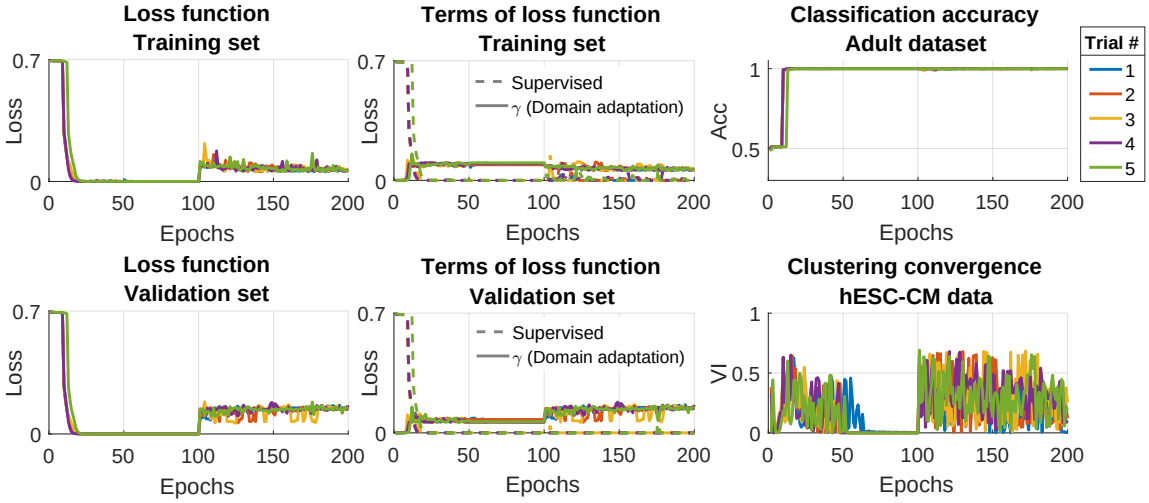


Figure 3.4: Training results of the DA-Sup-LSTM network.

The third column of Figure 3.4 shows that the classification accuracy in the adult domain stays close to one before and after incorporating the domain adaptation term. Although the clustering results had converged during initial-

ization stage (VI close to zero), they do not converge when the domain adaptation term is incorporated, which means that multiple sets of weights generate similar values in the loss function but different clustering results. As it can be seen in Figure 3.5, some variability is observed between trials in the set of weights at the last epoch of training.

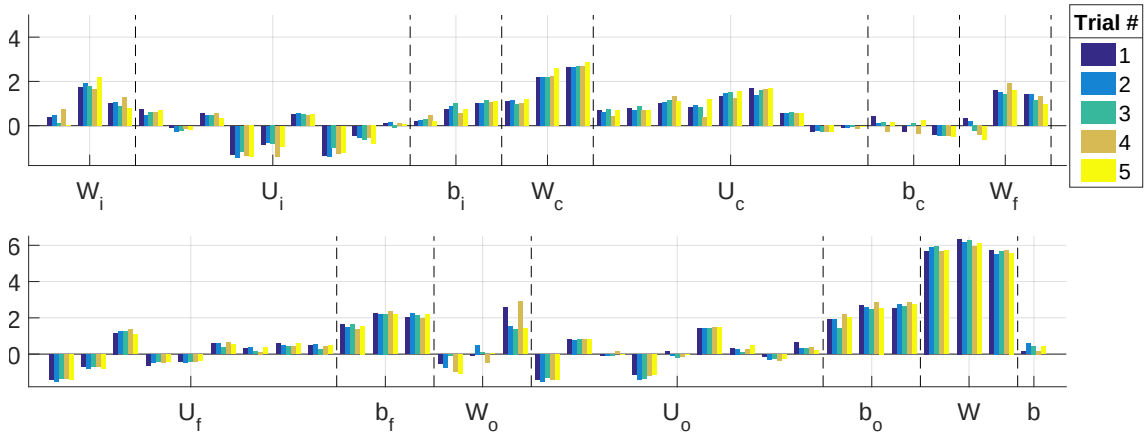
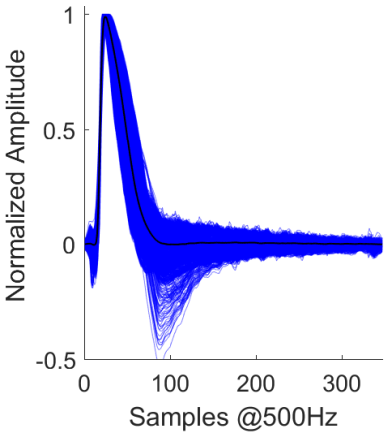


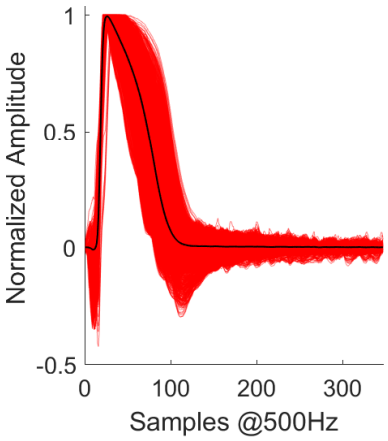
Figure 3.5: Weights of the DA-Sup-LSTM network at the last epoch of training for five different trials.

The *average prediction* of the network over trials was computed. Figure 3.6a and Figure 3.6b present the samples from the optical mapping dataset classified as atrial-like and ventricular-like, respectively. Figure 3.6c shows their distributions in the cell aggregates.

The *average prediction* of the network across trials was also computed for the single cell recording dataset. Figure 3.7 presents the histograms per class (according to ground truth labels). Observe that all the atrial-like samples are correctly classified ($\hat{y} < 0.5$) and only two ventricular-like samples are misclas-



(a) Atrial-like APs

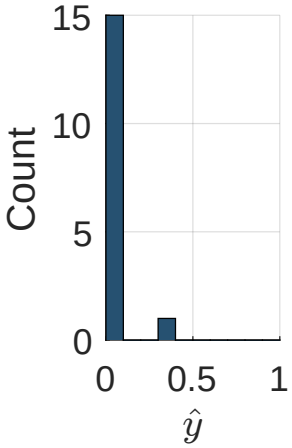


(b) Ventricular-like APs

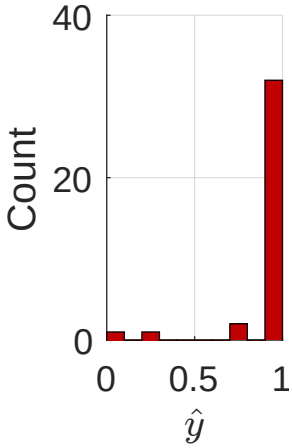
(c) Results in cell aggregates

Figure 3.6: Results of the DA-Sup-LSTM network in unlabeled hESC-CM dataset (DBI: 0.2258).

sified, which leads to 96.15% of classification accuracy.



(a) Atrial-like hESC-CMs



(b) Ventricular-like hESC-CMs

Figure 3.7: Results of the DA-Sup-LSTM network in labeled hESC-CM dataset (Classification accuracy: 0.9615).

Table 3.1 summarizes the performance of the domain adaptation approach on supervised learning for the individual trials and also for the *average prediction* across trials. The accuracy corresponds to the classification accuracy obtained in the single cell recording dataset at the last epoch, and the DBI corresponds to the clustering quality index computed on the optical mapping dataset at the last epoch of training.

	Trial #1	**Trial #2**	**Trial #3**	**Trial #4**	**Trial #5**	**Average prediction**
Accuracy	0.9423	0.9423	0.9423	0.9615	0.9615	0.9615
DBI	0.2260	0.2324	0.2269	0.2379	0.2262	0.2258

Table 3.1: Summary performance of the DA-Sup-LSTM network.

In Figure 3.8 one of the trials is taken as an example to illustrate the effect of domain adaptation in the distribution of the samples in feature space. Figure 3.8a shows the feature space $\varphi_{\mathcal{W}}(\mathrm{x})$ at the end of the initialization stage (Sup-LSTM), in which adult data is depicted as black circles, and embryonic data as yellow dots. 1600 adult samples have been plotted, but only two black circles are observed because all atrial samples are mapped close together, and ventricular samples are mapped close together but far from atrial samples. Therefore, the distribution of adult data in feature space can be modeled as the sum of two delta functions. Embryonic samples however seem to form a "path" in feature space that connects atrial and ventricular adult samples. Figure 3.8c shows the same data but in a 2D view. In order to estimate the density of em-

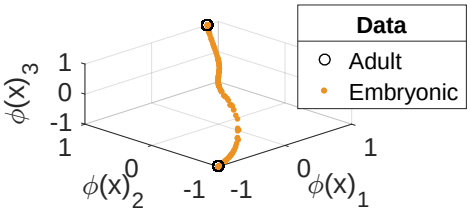
bryonic samples along this path we mapped the feature vector $\varphi_{\mathcal{W}}(\mathbf{x})$ to $\mathbb{R}$ by the parameters W and b of the output layer, and the histogram of $\varphi_{\mathcal{W}}(\mathbf{x})^T W + b$ is shown in Figure 3.8e. A magnified version of this histogram is presented in Figure 3.8g, where it can be easily observed that a significant amount of embryonic samples is mapped to a region of the space in which there are not adult samples.

The second column of Figure 3.8 depicts how the distribution of the data in feature space changes after domain adaptation. As it can be seen, although there are still some embryonic samples mapped somewhere in between, the distribution of embryonic samples approximates much better the distribution of samples observed in the adult domain. Consequently, the classification task is learned in regions of the feature space that not only represent the source domain, but also the target domain. This fact is precisely what allows the domain adaptation approach to perform much better than the supervised and semi-supervised approaches presented before.

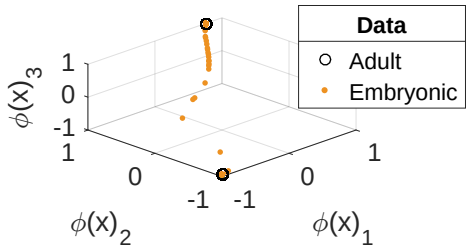
3.5.4.2 DA-Semi-LSTM-E

While in the domain adapted supervised approach (DA-Sup-LSTM) described in the previous section we trained the LSTM parameters by first running 100 epochs with $\gamma = 0$ and then 100 epochs with $\gamma = 10^4$, in the domain adapted

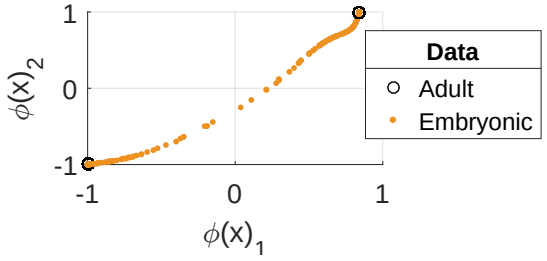
[4] Always using $\lambda = 0$ since it corresponds to the supervised case.



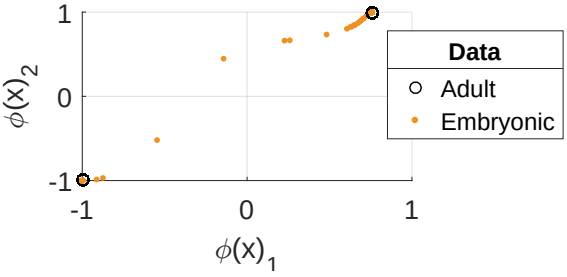
(a) Feature space, Epoch 100



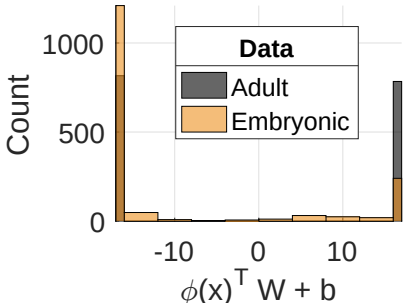
(b) Feature space, Epoch 200



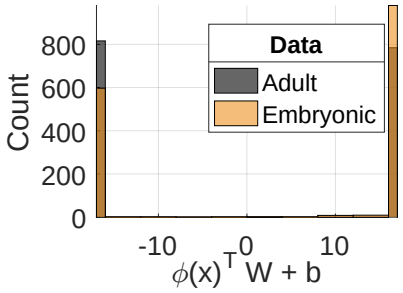
(c) Feature space, 2D view, Epoch 100



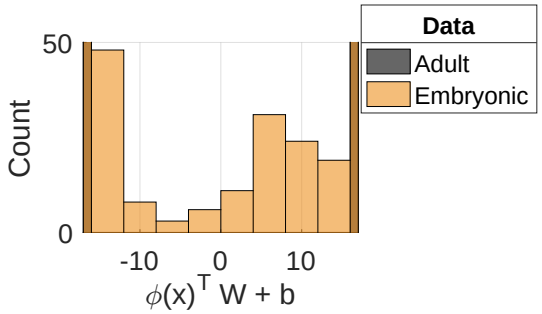
(d) Feature space, 2D view, Epoch 200



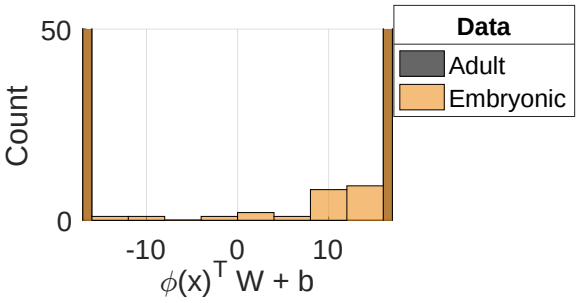
(e) Histogram $\varphi_{\mathcal{W}}(x)^T W + b$, Epoch 100



(f) Histogram $\varphi_{\mathcal{W}}(x)^T W + b$, Epoch 200



(g) Histogram $\varphi_{\mathcal{W}}(x)^T W + b$ (zoom in), Epoch 100



(h) Histogram $\varphi(x)^T W + b$ (zoom in), Epoch 200

Figure 3.8: Feature space representation before and after domain adaptation

semi-supervised approach with Euclidean distances (DA-Semi-LSTM-E) we obtained better results by training all LSTM parameters at once for 200 epochs with $\gamma = 1$ and $\lambda = 0.1$. Figure 3.9 presents the training results. The first and second columns show that the loss function as well as its semi-supervised and domain adaptation terms converge to similar values across trials in both the training (top) and validation (bottom) sets. The third column indicates that the network learns to correctly classify samples in the adult domain (top), and although the clustering results do not converge (bottom), the distance between clusterings in consecutive epochs (VI) is significantly reduced by the end of training. The sets of weights at the last epoch of training for each trial are shown in Figure 3.10, where some variability is observed.

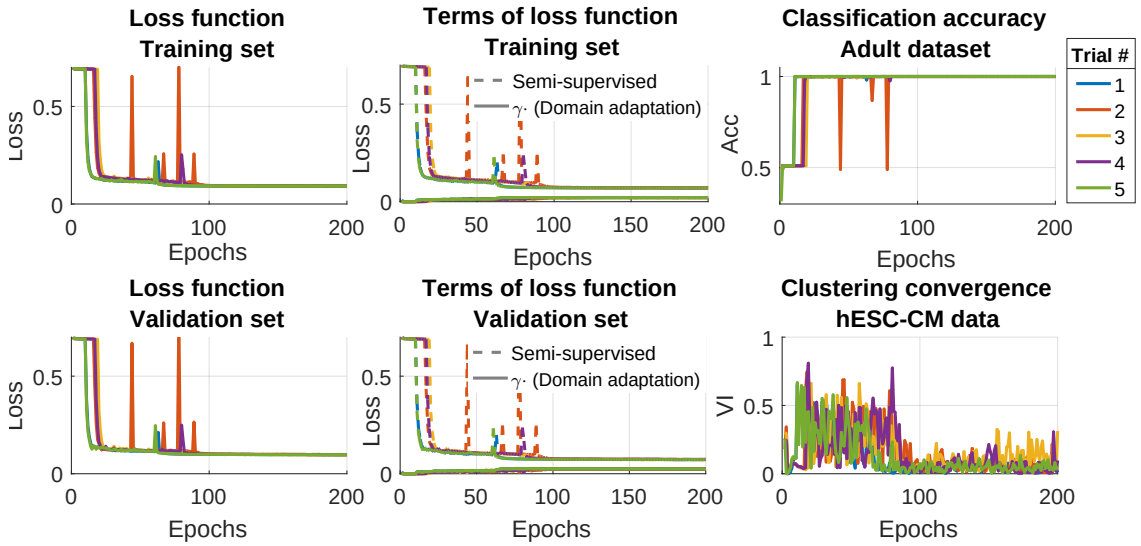


Figure 3.9: Training results of the DA-Semi-LSTM-E network.

The *average prediction* of the DA-Semi-LSTM-E network across trials per

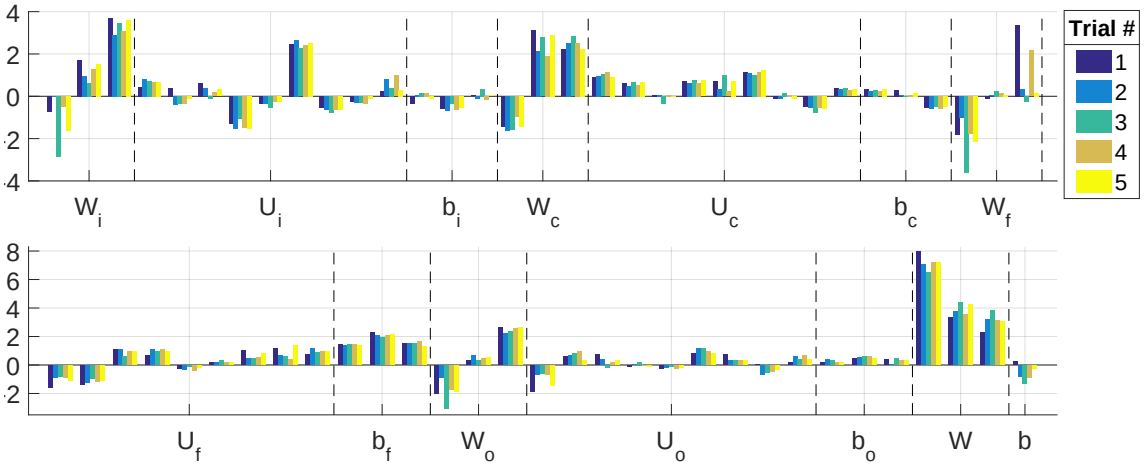


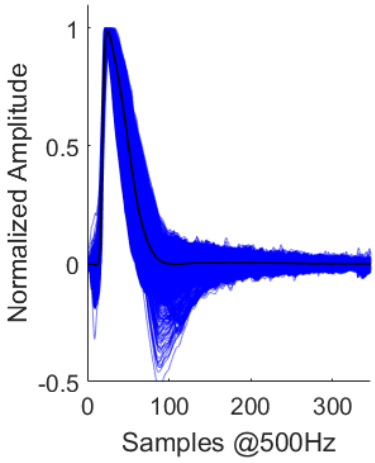
Figure 3.10: Weights of the DA-Semi-LSTM-E network at the last epoch of training for five different trials.

sample was computed. Figure 3.11a shows the samples classified as atrial-like in the optical mapping dataset, and Figure 3.11b shows the ones classified as ventricular-like. Figure 3.11c depicts their distribution on the cell clusters. This method achieves a DBI of 0.2304 in the *average prediction* [5].

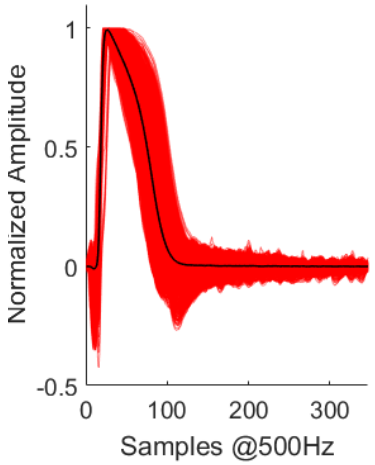
Figure 3.12 shows the histograms of the *average prediction* of the network across trials on the single cell recording dataset, organized according to their ground truth labels. As it can be noted, all the atrial-like samples are correctly classified ($\hat{y} < 0.5$), and only 5 ventricular-like samples are misclassified, reaching a classification accuracy of 90.38%.

Table 3.2 presents the summary of the performance of the DA-Semi-LSTM-E network for the individual trials and for the *average prediction*. Observe that

[5]The *average prediction* for a given sample is computed by averaging the output of the network across trials at their last epoch of training. Then the DBI is computed on the clustering results generated by the *average prediction* in the optical mapping dataset.



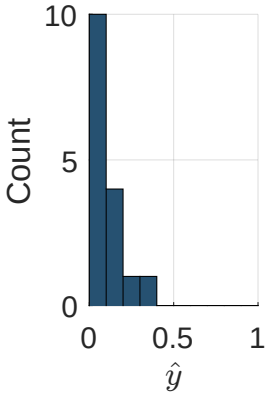
(a) Atrial-like APs



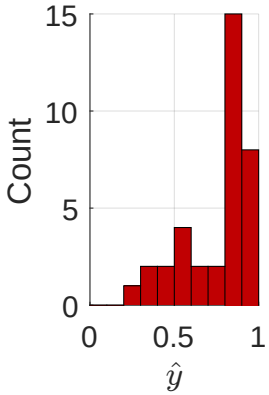
(b) Ventricular-like APs

(c) Results in cell aggregates

Figure 3.11: Results of the DA-Semi-LSTM-E network in unlabeled hESC-CM dataset (DBI: 0.2304).



(a) Atrial-like hESC-CMs



(b) Ventricular-like hESC-CMs

Figure 3.12: Results of the DA-Semi-LSTM-E network in labeled hESC-CM dataset (Classification accuracy: 0.9038).

in this case there exists significant variability of performance between trials: while the DA-Semi-LSTM-E method achieves the best performance in Trial #5 with 96.15% of accuracy in the single cell recording dataset (only 2 misclassified samples), in Trial #4 it obtains only 76.92% of accuracy (12 misclassified samples).

	Trial #1	**Trial #2**	**Trial #3**	**Trial #4**	**Trial #5**	**Average prediction**
Accuracy	0.9038	0.8654	0.9231	0.7692	0.9615	0.9038
DBI	0.2288	0.2304	0.2294	0.2396	0.2258	0.2304

Table 3.2: Summary performance of the DA-Semi-LSTM-E network.

3.5.4.3 DA-Semi-LSTM-M

In this case the network is trained in a two step process: first minimize the supervised term with $\gamma = 0$ and $\lambda = 0$ for 100 epochs, and then minimize the unsupervised and domain adaptation terms with $\lambda = 0.1$ and $\gamma = 5$, also for 100 epochs. The first column of Figure 3.13 shows that the loss function exhibits abrupt changes due to the two step training and that all trials follow similar trajectories despite some oscillations. Again a slightly increasing trend is noted in the loss function evaluated in the validation set, which can be indicative of overfitting.

The abrupt increase of the semi-supervised term observed in the second column of Figure 3.13 is explained by the incorporation of the unsupervised term

(before epoch 100 only the supervised part is plotted). The domain adaptation term is always plotted (although it is not being optimized in the first 100 epochs), and its small increase at epoch 100 is explained because there is a compromise between optimization of the unsupervised and domain adaptation terms.

In the third column of Figure 3.13 (top), although the classification of adult samples is not perfect across the epochs, all trials ultimately re-learn to classify adult samples after unsupervised and domain adaptation terms are incorporated.

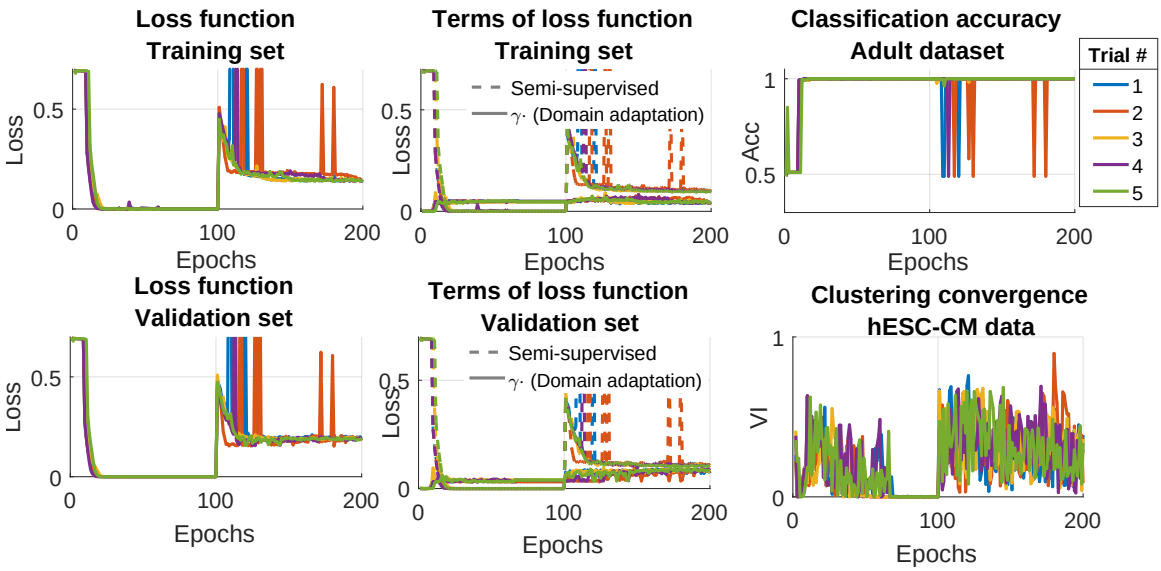


Figure 3.13: Training results of the DA-Semi-LSTM-M network.

Figure 3.14 depicts the set of weights at the last epoch of training in different trials. As observed, there exists significant variability. Part of the variability can be related to the oscillating behavior observed in the VI (third column

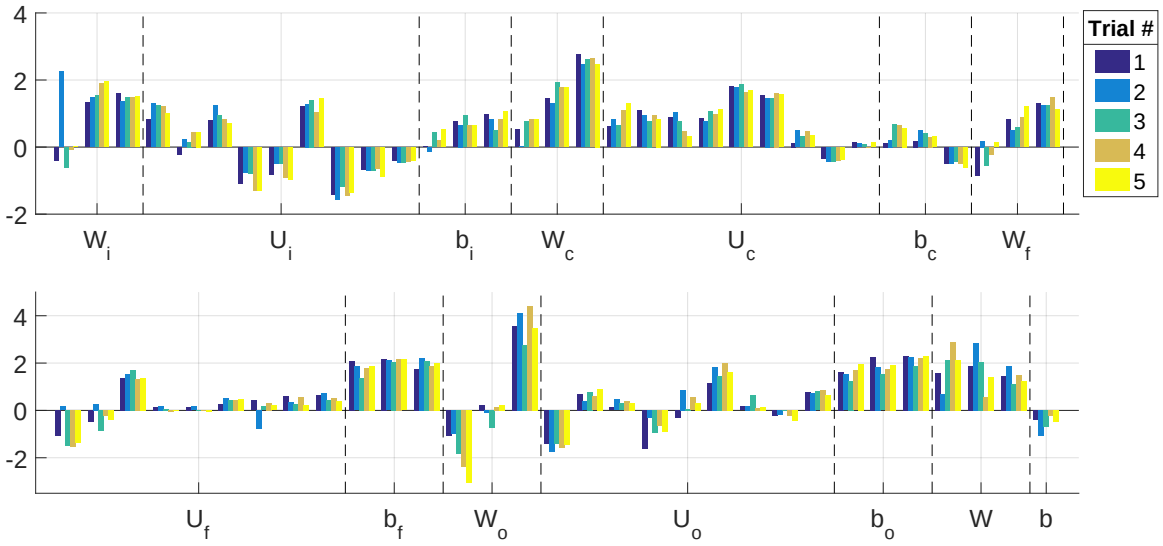
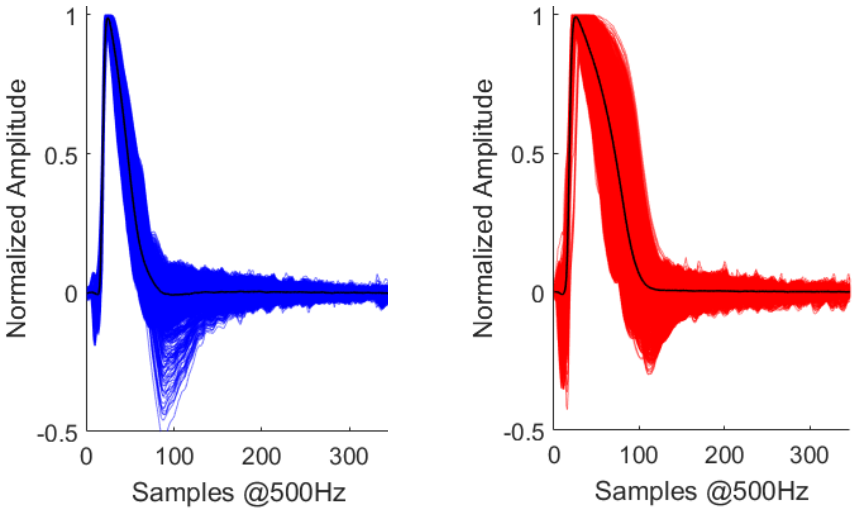


Figure 3.14: Weights of the DA-Semi-LSTM-M network at the last epoch of training for five different trials.

of Figure 3.13 (bottom)), showing that there is not convergence in the clustering results. Additional epochs of training might help to reduce this variability since there is a decreasing trend in the VI that does not seem to have converged by the end of epoch #200.

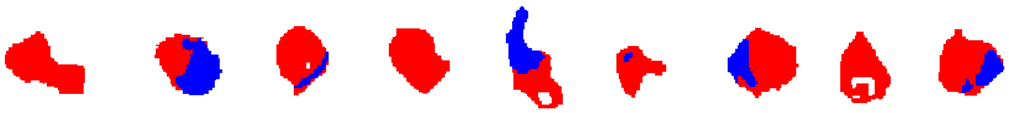
The *average prediction* of the DA-Semi-LSTM-M network across trials was computed. Figure 3.15a and Figure 3.15b show the samples classified as atrial-like and ventricular-like in the optical mapping dataset, respectively. Figure 3.15c depicts the spatial distribution of the classes in the cell aggregates. DA-Semi-LSTM-M generates a DBI of 0.2188, which exceeds the performance of all the alternative approaches.

Regarding the single cell recording dataset, Figure 3.16 shows the histograms of the average output per sample, organized according to their ground truth la-



(a) Atrial-like APs

(b) Ventricular-like APs

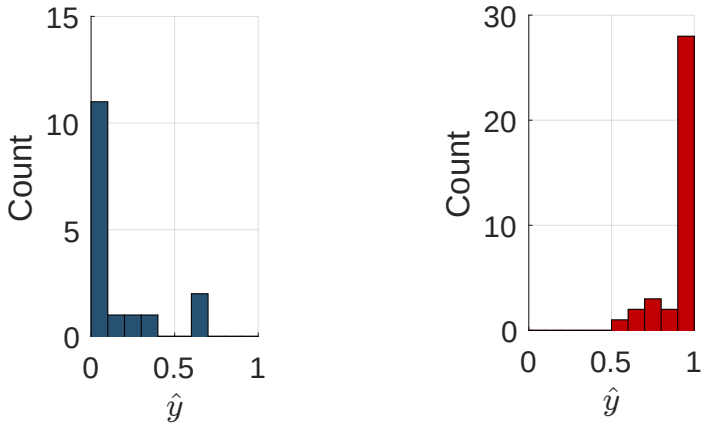


(c) Results in cell aggregates

Figure 3.15: Results of the DA-Semi-LSTM-M network in unlabeled hESC-CM dataset (DBI: 0.2188).

bels. As observed, all ventricular-like samples are correctly classified, while 2 atrial-like samples are misclassified. Therefore, the DA-Semi-LSTM-M network achieves a classification accuracy of 96.15%.

Table 3.3 shows the performance of DA-Semi-LSTM-M network for each trial and also for the *average prediction* of the network across trials. The accuracy corresponds to the classification accuracy in the single-cell recording dataset, and DBI corresponds to the clustering quality in the optical mapping dataset. As observed, DA-Semi-LSTM-M method reaches 100% of accuracy on the single cell recording dataset in Trial #5, but it performs better in terms of



(a) Atrial-like hESC-CMs

(b) Ventricular-like hESC-CMs

Figure 3.16: Results of the DA-Semi-LSTM-M network in labeled hESC-CM dataset (Classification accuracy: 0.9615).

clustering quality in Trial #1 (lowest DBI).

	Trial #1	**Trial #2**	**Trial #3**	**Trial #4**	**Trial #5**	**Average prediction**
Accuracy	0.9231	0.9231	0.9231	0.9231	1.0	0.9615
DBI	0.2100	0.2438	0.2185	0.2242	0.2226	0.2188

Table 3.3: Summary performance of the DA-Semi-LSTM-M network.

3.5.5 Analysis

Experimental results show that the minimization of the MMD statistic between embryonic and adult samples in the proposed training scheme leads to more similar distributions of embryonic and adult data in feature space (see Figure 3.8). Moreover, the implementation of domain adaptation via minimization of MMD proves to be a powerful tool for the classification of hESC-CM

APs: it allows for significant improvements in the performance of the three approaches presented in Chapter 2 by simply incorporating an additional term to the loss function. In the supervised learning case, the classification accuracy in the single cell recording dataset improved from 32.69% (Sup-LSTM) to 96.15% (DA-Sup-LSTM), and the DBI decreased from 0.2834 to 0.2258 in the optical mapping dataset. In the semi-supervised learning case with Euclidean distances there was an improvement from 76.92% (Semi-LSTM-E) to 90.38% (DA-Semi-LSTM-E) in classification accuracy and a decrease in DBI from 0.2458 to 0.2341. Moreover, the classification accuracy improved from 76.92% (Semi-LSTM-M) to 96.15% (DA-Semi-LSTM-M) and the DBI decreased from 0.2390 to 0.2188 in the semi-supervised learning case with metamorphosis distances[6].

Although the slightly increasing trend consistently observed in the domain adaptation term evaluated in the validation set could be indicative of overfitting (see Figures 3.4 and 3.13), this is contradicted by the high level of performance achieved by the domain adaptation approach when tested in a completely different dataset (single cell recording dataset). Therefore, the slightly increasing trend is attributable to the particular subset of data chosen for validation, which could be further analyzed by cross-validation experiments.

Figure 3.17 presents the results of the proposed domain adapted networks along with the results of the methods proposed in [1] (1NN method with Eu-

[6]Figure 4.1 shows a comparison of all these cases.

Method	1NN	1NN	DA-Sup-LSTM	DA-Semi-LSTM	DA-Semi-LSTM
Templates	20 [1]	20 [1]	800 SMRS	800 SMRS	800 SMRS
Metric	M	E		E	M
Accuracy	**0.9615**	0.8654	**0.9615**	0.9038	**0.9615**
DBI	0.2297	0.2558	0.2258	0.2341	**0.2188**

Table 3.4: Comparing the results of the proposed domain adaptation method with the results presented in [1]. Accuracy is computed in single cell recording hESC-CM dataset. DBI is computed in optical mapping dataset. (E: Euclidean, M: Metamorphosis).

clidean and metamorphosis distances). It can be observed that the *average prediction* of the domain adapted networks is visually similar to the 1NN metamorphosis classification in the three cases (DA-Sup-LSTM, DA-Semi-LSTM-E, and DA-Semi-LSTM-M). Table 3.4 and Figure 3.18 further compare the performance of the proposed domain adaptation method with the state of the art for this problem (1NN with metamorphosis distances) in terms of classification accuracy in the single cell recording dataset and clustering quality (DBI) in the optical mapping dataset. Interestingly, the addition of the domain adaptation term to the supervised loss is enough to not only match the accuracy of the state of the art (96.15%), but also outperform it in terms of clustering quality. This is a powerful result since DA-Sup-LSTM approach also has significant computational advantages with respect to the state of the art: it does not require any metamorphosis computation, and the classification of a new sample simply corresponds to a forward pass through a small RNN with fixed weights (as opposed to solving 20 computational expensive optimization problems as the state-of-the-art method requires).

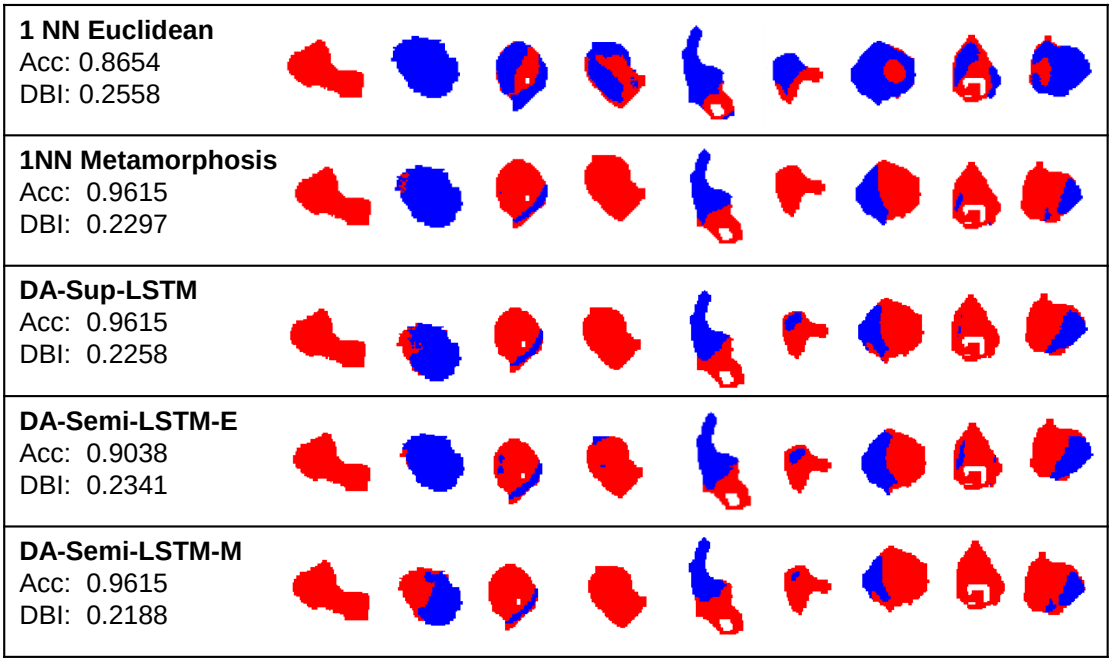


Figure 3.17: Comparison of the results of domain adaptation approach with the results presented in [1]. Acc corresponds to the classification accuracy in the single cell recording dataset, and DBI is the clustering quality index in the optical mapping dataset.

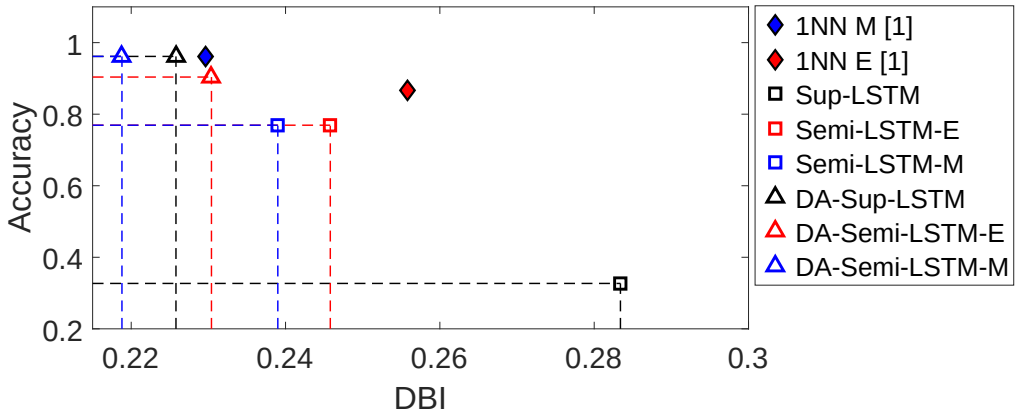


Figure 3.18: Accuracy in single cell recording dataset vs DBI in optical mapping dataset. (E: Euclidean, M: Metamorphosis).

The addition of the unsupervised term with metamorphosis distances (DA-Semi-LSTM-M) improves the clustering quality even further and still preserves

the high accuracy in the labeled dataset. However, in the DA-Semi-LSTM-M network the decrease in the DBI (from 0.2258 to 0.2188) comes at the cost of a time-consuming preprocessing phase in which 1500 metamorphosis distances need to be computed. As it can be seen in Figure 3.18, the DA-Semi-LSTM-E does not perform better than the state of the art, but it outperforms both 1NN with Euclidean distances, Semi-LSTM-E, and Semi-LSTM-M approaches in terms of accuracy and clustering quality. The fact that DA-Sup-LSTM performs better than DA-Semi-LSTM-E would suggest that the unsupervised term with Euclidean distances does not provide meaningful information for the task; however additional experiments exhaustively exploring the parameter space (λ, γ) would be required to support such a conclusion.

3.6 Chapter summary

In this chapter we have presented a domain adaptation approach to classification of hESC-CMs that builds on the supervised and semi-supervised learning approaches presented in Chapter 2. Domain adaptation is implemented by incorporating an estimator of the MMD statistic into the loss function to train an RNN with LSTM units. Experimental results show that the proposed domain adaptation approach not only outperforms the state-of-the-art method for this problem, but also presents significant computational advantages.

Chapter 4

Conclusions

In this thesis we have proposed a recurrent neural network with LSTM units for classification of hESC-CM APs and we have presented two different methods for training it: semi-supervised learning and domain adapted learning. Both of them showed significant computational advantages with respect to the state of the art in the classification of new samples. While the state-of-the-art 1NN metamorphosis method requires the solution of 20 computational intensive optimization problems, our approach just requires a forward pass through a small RNN with fixed weights. In this regard, one of the main drawbacks of existing approaches to classification of hESC-CM APs is their inability to efficiently scale up to large datasets, which is fairly simple in the case of our RNN classifier. Figure 4.1 presents a summary of our results compared to those presented in [1] for 1NN classification with Euclidean and metamorphosis dis-

tances in terms of accuracy in the single cell recording dataset, and DBI and spatial distribution of classes in the optical mapping dataset.

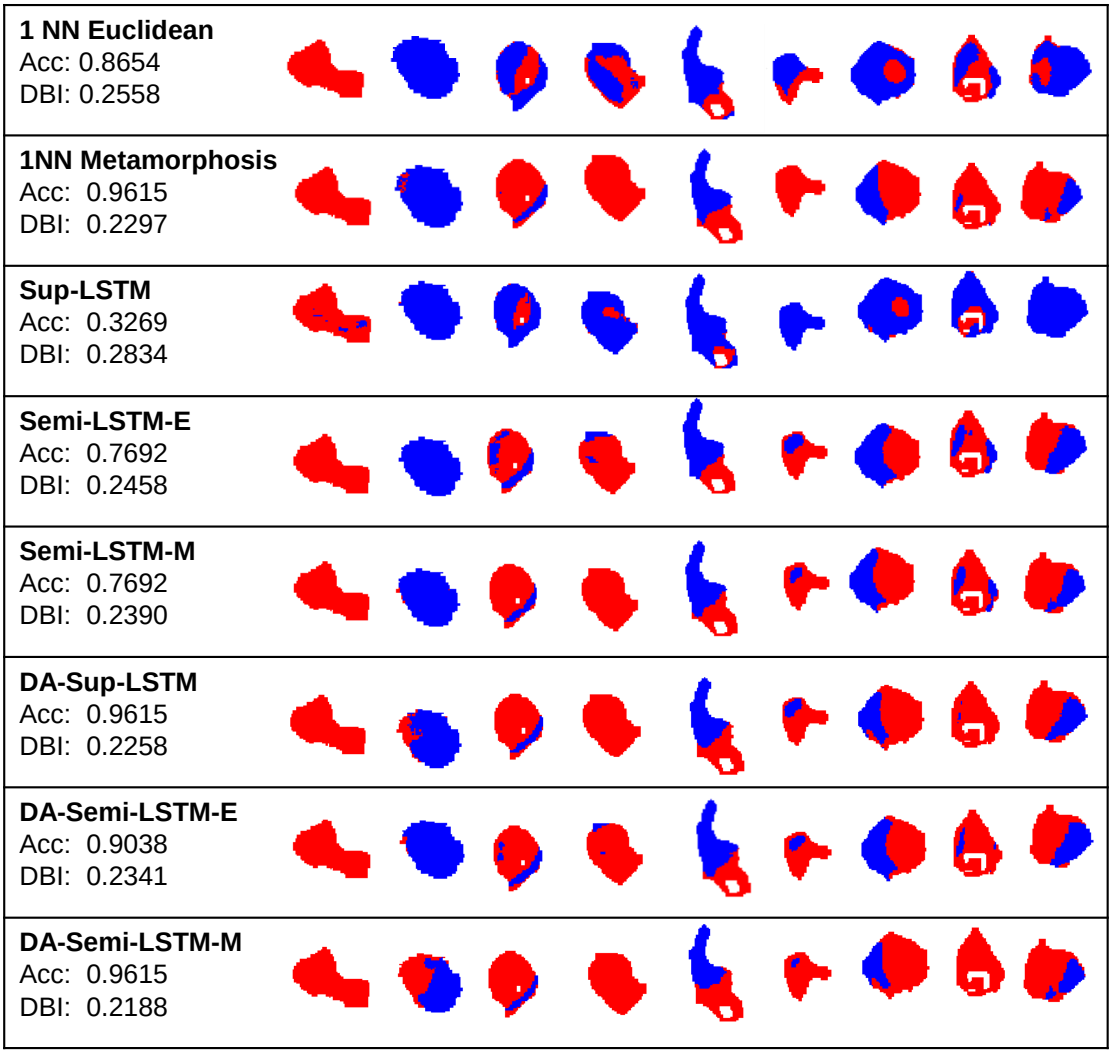


Figure 4.1: Summary of results. Acc corresponds to the classification accuracy in the single cell recording dataset, and DBI is the clustering quality index in the optical mapping dataset. DA-Semi-LSTM-M obtains the best performance (lowest DBI and 96.15% classification accuracy). 1NN Metamorphosis and DA-Sup-LSTM also achieve 96.15% of accuracy, but present higher DBI.

Compared to the 1NN method with Euclidean distances, the proposed semi-

supervised learning scheme proves to be a more suitable way of using Euclidean distances. It leverages the fact that the Euclidean metric is a good approximation of metamorphosis when distances are small. Therefore, computing Euclidean distances within the embryonic domain in terms of similarity factors provides more accurate information than computing Euclidean distances between different domains. However, this seems to be valid only when embryonic test data are similar to embryonic training data: the proposed semi-supervised learning approach does not exhibit great transferability capacity, since it generates lower classification accuracy than the 1NN approach when tested on a different embryonic dataset.

One of the important challenges in the classification of hESC-CMs is the computational complexity of the metamorphosis algorithm. That is why we implemented a modified mini-batch SGD optimization algorithm that uses predefined random mini-batches (which allows us to train the network based on a limited number of metamorphosis distances computed before training). We experimentally analyzed the effect of this modification, showing that it has the potential to behave similar to classic mini-batch SGD.

The poor performance achieved by a baseline supervised learning method trained only on adult CM data confirms the fact that adult and hESC-CM APs belong to different domains. This is further supported by the significant improvement observed with the proposed domain adapted learning scheme. The

implementation of domain adaptation via minimization of MMD in the training of the RNN classifier not only is more computational efficient than the state of the art but also outperforms it in terms of clustering quality. Moreover, it generates high classification accuracy in a completely different dataset without retraining, which demonstrates its transfer learning abilities.

We have shown that the integration of domain adaptation in supervised and semi-supervised approaches surpasses the state of the art for this problem. However, there are many ways in which this framework can be further exploited. First of all, we restricted the problem to binary classification since, to the best of our knowledge, computational models of human pacemaker cells are not as validated as models of atrial and ventricular cells [64]. Nevertheless, from a mathematical perspective the number of phenotypes is not a constraint, so if reliable adult pacemaker APs are available, minimum modifications are required to include them in this framework. Second, although our results satisfactorily demonstrated the potential of the proposed approaches, an exhaustive exploration of the space of hyperparameters (mainly λ and γ but also n_e, n_a, ϵ, among others) is needed to find its optimal performance. Lastly, the results presented in this thesis are significantly limited by the use of indirect metrics such as a clustering quality index (DBI), which is forced by the absence of a large ground truth labeled embryonic dataset. Therefore, an important step to further validate the proposed approaches would be to evaluate its classification

accuracy in a large and reliably labeled hESC-CM APs dataset.

Bibliography

[1] G. Gorospe, R. Zhu, J.-Q. He, L. Tung, L. Younes, and R. Vidal, "Efficient metamorphosis computation for classifying embryonic cardiac action potentials," in *5th Workshop on Mathematical Foundations of Computational Anatomy*, 2015.

[2] World Health Organization. (2018) Global health observatory data: Top 10 causes of death. [Online]. Available: http://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/

[3] ——. (2017) World heart day. [Online]. Available: http://www.who.int/cardiovascular_diseases/world-heart-day-2017/en/

[4] S. L. Beeres, D. E. Atsma, J. van Ramshorst, M. J. Schalij, and J. J. Bax, "Cell therapy for ischaemic heart disease," *Heart*, vol. 94, no. 9, pp. 1214–1226, 2008.

[5] M. E. Hartman, J. J. Chong, and M. A. Laflamme, "State of the art in car-

diomyocyte transplantation," in *Cardiac Regeneration*. Springer, 2017, pp. 177–218.

[6] M. A. Laflamme, K. Y. Chen, A. V. Naumova, V. Muskheli, J. A. Fugate, S. K. Dupras, H. Reinecke, C. Xu, M. Hassanipour, S. Police *et al.*, "Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts," *Nature Biotechnology*, vol. 25, no. 9, p. 1015, 2007.

[7] O. Caspi, I. Huber, I. Kehat, M. Habib, G. Arbel, A. Gepstein, L. Yankelson, D. Aronson, R. Beyar, and L. Gepstein, "Transplantation of human embryonic stem cell-derived cardiomyocytes improves myocardial performance in infarcted rat hearts," *Journal of the American College of Cardiology*, vol. 50, no. 19, pp. 1884–1893, 2007.

[8] J. J. Chong, X. Yang, C. W. Don, E. Minami, Y.-W. Liu, J. J. Weyers, W. M. Mahoney, B. Van Biber, S. M. Cook, N. J. Palpant *et al.*, "Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts," *Nature*, vol. 510, no. 7504, p. 273, 2014.

[9] I. Kehat, D. Kenyagin-Karsenti, M. Snir, H. Segev, M. Amit, A. Gepstein, E. Livne, O. Binah, J. Itskovitz-Eldor, and L. Gepstein, "Human embryonic stem cells can differentiate into myocytes with structural and functional

properties of cardiomyocytes," *The Journal of Clinical Investigation*, vol. 108, no. 3, pp. 407–414, 2001.

[10] C. Xu, S. Police, N. Rao, and M. K. Carpenter, "Characterization and enrichment of cardiomyocytes derived from human embryonic stem cells," *Circulation Research*, vol. 91, no. 6, pp. 501–508, 2002.

[11] K. Turksen, *Human embryonic stem cell protocols*. Springer, 2006, vol. 331.

[12] W.-Z. Zhu, K. D. Hauch, C. Xu, and M. A. Laflamme, "Human embryonic stem cells and cardiac repair," *Transplantation Reviews*, vol. 23, no. 1, pp. 53–68, 2009.

[13] O. Caspi, I. Itzhaki, I. Kehat, A. Gepstein, G. Arbel, I. Huber, J. Satin, and L. Gepstein, "In vitro electrophysiological drug testing using human embryonic stem cell derived cardiomyocytes," *Stem Cells and Development*, vol. 18, no. 1, pp. 161–172, 2009.

[14] S. R. Braam, L. Tertoolen, A. van de Stolpe, T. Meyer, R. Passier, and C. L. Mummery, "Prediction of drug-induced cardiotoxicity using human embryonic stem cell-derived cardiomyocytes," *Stem Cell Research*, vol. 4, no. 2, pp. 107–116, 2010.

[15] A. D. Ebert and C. N. Svendsen, "Human stem cells and drug screen-

ing: opportunities and challenges," *Nature Reviews Drug Discovery*, vol. 9, no. 5, p. 367, 2010.

[16] S. Peng, A. E. Lacerda, G. E. Kirsch, A. M. Brown, and A. Bruening-Wright, "The action potential and comparative pharmacology of stem cell-derived human cardiomyocytes," *Journal of Pharmacological and Toxicological Methods*, vol. 61, no. 3, pp. 277–286, 2010.

[17] J.-Q. He, Y. Ma, Y. Lee, J. A. Thomson, and T. J. Kamp, "Human embryonic stem cells develop into multiple types of cardiac myocytes: action potential characterization," *Circulation Research*, vol. 93, no. 1, pp. 32–39, 2003.

[18] L. Sartiani, E. Bettiol, F. Stillitano, A. Mugelli, E. Cerbai, and M. E. Jaconi, "Developmental changes in cardiomyocytes differentiated from human embryonic stem cells: a molecular and electrophysiological approach," *Stem Cells*, vol. 25, no. 5, pp. 1136–1144, 2007.

[19] A. Nygren, C. Fiset, L. Firek, J. Clark, D. Lindblad, R. Clark, and W. Giles, "Mathematical model of an adult human atrial cell: the role of k+ currents in repolarization," *Circulation Research*, vol. 82, no. 1, pp. 63–81, 1998.

[20] T. O'Hara, L. Virág, A. Varró, and Y. Rudy, "Simulation of the undiseased human cardiac ventricular action potential: model formulation and experimental validation," *PLoS Computational Biology*, vol. 7, no. 5, p. e1002061, 2011.

[21] M. L. Vestergaard, S. Grubb, K. Koefoed, Z. Anderson-Jenkins, K. Grunnet-Lauridsen, K. Calloe, C. Clausen, S. T. Christensen, K. Møllgard, and C. Y. Andersen, "Human embryonic stem cell-derived cardiomyocytes self-arrange with areas of different subtypes during differentiation," *Stem Cells and Development*, vol. 26, no. 21, pp. 1566–1577, 2017.

[22] M. M. Maleckar, J. L. Greenstein, W. R. Giles, and N. A. Trayanova, "K+ current changes account for the rate dependence of the action potential in the human atrial myocyte," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 297, no. 4, pp. H1398–H1410, 2009.

[23] K. H. Ten Tusscher and A. V. Panfilov, "Alternans and spiral breakup in a human ventricular tissue model," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 291, no. 3, pp. H1088–H1100, 2006.

[24] E. Grandi, F. S. Pasqualini, and D. M. Bers, "A novel computational model of the human ventricular action potential and ca transient," *Journal of Molecular and Cellular Cardiology*, vol. 48, no. 1, pp. 112–121, 2010.

[25] A. Fabbri, M. Fantini, R. Wilders, and S. Severi, "Computational analysis of the human sinus node action potential: model development and effects of mutations," *The Journal of Physiology*, vol. 595, no. 7, pp. 2365–2396, 2017.

[26] S. Weinberg, E. A. Lipke, and L. Tung, "In vitro electrophysiological mapping of stem cells," in *Stem Cells for Myocardial Regeneration*. Springer, 2010, pp. 215–237.

[27] R. Zhu, M. A. Millrod, E. T. Zambidis, and L. Tung, "Variability of action potentials within and among cardiac cell clusters derived from human embryonic stem cells," *Scientific Reports*, vol. 6, p. 18544, 2016.

[28] D. T. Du, N. Hellen, C. Kane, and C. M. Terracciano, "Action potential morphology of human induced pluripotent stem cell-derived cardiomyocytes does not predict cardiac chamber specificity and is dependent on cell density," *Biophysical Journal*, vol. 108, no. 1, pp. 1–4, 2015.

[29] W. R. Giles and D. Noble, "Rigorous phenotyping of cardiac ipsc preparations requires knowledge of their resting potential (s)," *Biophysical Journal*, vol. 110, no. 1, p. 278, 2016.

[30] C. Kane, D. T. Du, N. Hellen, and C. M. Terracciano, "The fallacy of assigning chamber specificity to ipsc cardiac myocytes from action potential morphology," *Biophysical Journal*, vol. 110, no. 1, p. 281, 2016.

[31] G. C. Bett, A. D. Kaplan, and R. L. Rasmusson, "Action potential shape is a crucial measure of cell type of stem cell-derived cardiocytes," *Biophysical Journal*, vol. 110, no. 1, p. 284, 2016.

[32] A. Muszkiewicz, O. J. Britton, P. Gemmell, E. Passini, C. Sánchez, X. Zhou, A. Carusi, T. A. Quinn, K. Burrage, A. Bueno-Orovio *et al.*, "Variability in cardiac electrophysiology: using experimentally-calibrated populations of models to move beyond the single virtual physiological human paradigm," *Progress in Biophysics and Molecular Biology*, vol. 120, no. 1-3, pp. 115–127, 2016.

[33] A. Margolis, "A literature review of domain adaptation with unlabeled data," *Technical Report*, pp. 1–42, 2011.

[34] G. Gorospe, L. Younes, L. Tung, and R. Vidal, "A metamorphosis distance for embryonic cardiac action potential interpolation and classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 469–476.

[35] A. Trouvé and L. Younes, "Metamorphoses through lie group action," *Foundations of Computational Mathematics*, vol. 5, no. 2, pp. 173–198, 2005.

[36] G. Gorospe, R. Zhu, M. A. Millrod, E. T. Zambidis, L. Tung, and R. Vidal, "Automated grouping of action potentials of human embryonic stem cell-derived cardiomyocytes," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 9, pp. 2389–2395, 2014.

[37] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[38] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *International Conference on Machine Learning*, 2014, pp. 1764–1772.

[39] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2625–2634.

[40] S. Hochreiter and J. Schmidhuber, "Lstm can solve hard long time lag problems," in *Advances in Neural Information Processing Systems*, 1997, pp. 473–479.

[41] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," *arXiv preprint arXiv:1710.06122*, 2017.

[42] C. Zhang, G. Wang, J. Zhao, P. Gao, J. Lin, and H. Yang, "Patient-specific ecg classification based on recurrent neural networks and clustering technique," in *2017 13th IASTED International Conference on Biomedical Engineering (BioMed)*, Feb 2017, pp. 63–67.

[43] P. Warrick and M. N. Homsi, "Cardiac arrhythmia detection from ecg combining convolutional and long short-term memory networks," in *2017 Computing in Cardiology (CinC)*, Sept 2017, pp. 1–4.

[44] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *Proceedings*. Presses universitaires de Louvain, 2015, pp. 89 – 94.

[45] S. Chauhan and L. Vig, "Anomaly detection in ecg time signals via deep long short-term memory networks," in *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Oct 2015, pp. 1–7.

[46] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two-sample-problem," in *Advances in Neural Information Processing Systems*, 2007, pp. 513–520.

[47] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017.

[48] F. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with LSTM." *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, 2000.

[49] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learn-

ing an invariant mapping," in *Computer Vision and Pattern Recognition, 2006 IEEE computer society conference on*, vol. 2. IEEE, 2006, pp. 1735–1742.

[50] F. Ratle, G. Camps-Valls, and J. Weston, "Semisupervised neural networks for efficient hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2271–2282, 2010.

[51] Y. Li, X. Tian, X. Shen, and D. Tao, "Classification and representation joint learning via deep networks," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press, 2017, pp. 2215–2221.

[52] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 2, pp. 224–227, 1979.

[53] M. Meilă, "Comparing clusterings by the variation of information," in *Learning Theory and Kernel Machines*. Springer, 2003, pp. 173–187.

[54] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1600–1607.

[55] V. Jacquemet, "Steady-state solutions in mathematical models of atrial

cell electrophysiology and their stability," *Mathematical Biosciences*, vol. 208, no. 1, pp. 241–269, 2007.

[56] S. Iravanian and L. Tung, "A novel algorithm for cardiac biosignal filtering based on filtered residue method," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 11, pp. 1310–1317, 2002.

[57] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[58] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[59] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[60] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction." in *AAAI*, vol. 8, 2008, pp. 677–682.

[61] L. Duan, I. W. Tsang, D. Xu, and S. J. Maybank, "Domain transfer svm for video concept detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 1375–1381.

[62] G. Dziugaite, D. Roy, and Z. Ghahramani, "Training generative neural networks via maximum mean discrepancy optimization," in *Uncertainty*

in Artificial Intelligence-Proceedings of the 31st Conference, UAI 2015, 2015, pp. 258–267.

[63] H.-Y. Chen and J.-T. Chien, "Deep semi-supervised learning for domain adaptation," in *Machine Learning for Signal Processing (MLSP), 2015 IEEE 25th International Workshop on*. IEEE, 2015, pp. 1–6.

[64] V. A. Maltsev, Y. Yaniv, A. V. Maltsev, M. D. Stern, and E. G. Lakatta, "Modern perspectives on numerical modeling of cardiac pacemaker cell," *Journal of Pharmacological Sciences*, vol. 125, no. 1, pp. 6–38, 2014.

Vita

Carolina Pacheco Oñate received the Electrical Engineering degree from University of Chile in 2014, where she worked in the design of fractional-order observers for integer-order systems as her final undergraduate research project. In 2016 she was awarded a Chilean scholarship to pursue the MSE on Biomedical Engineering (BME) at Johns Hopkins University (JHU).

She joined the "Computer Vision, Learning and Dynamics lab" at JHU in January 2017 to investigate the classification of hESC-CMs from the neural network perspective, incorporating concepts of semi-supervised learning and domain adaptation. Part of her work was selected for the 21st International Conference on Medical Image Computing and Computed Assisted Intervention (MICCAI 2018). She will continue her research at JHU as a BME PhD student.