Comparison of statistical, machine learning, and mathematical modelling methods to investigate the effect of ageing on dog's cardiovascular system

Elham Ataei, Sara Costa Faya, Haibo Liu, Damiano Lombardi, Sylvain Bernasconi, Pieter-Jan Guns, Michael Markert

ESAIM: PROCEEDINGS AND SURVEYS, Vol. ?, 2023, 1-10
Editors: Will be set by the publisher

COMPARISON OF STATISTICAL, MACHINE LEARNING, AND MATHEMATICAL MODELLING METHODS TO INVESTIGATE THE EFFECT OF AGEING ON DOG'S CARDIOVASCULAR SYSTEM

Elham Ataei Alizadeh[1], Sara Costa Faya[2], Haibo Liu[3,2], Damiano Lombardi[2], Sylvain Bernasconi[3], Pieter-Jan Guns[4] and Michael Markert[1]

Abstract. The aim of this work is to provide a preliminary comparison of different classes of methods to automatically detect the effect of ageing from *in vivo* data. The application which motivated this work is related to safety pharmacology, whose major goal is to determine, in a pre-clinical phase, whether a drug is potentially dangerous for the health [1]. In particular, we are going to compare statistical, machine learning and mathematical modelling methods.

Résumé. L'objectif de ce travail est de fournir une comparaison préliminaire entre différents classes de méthodes pour la détection automatique de l'effet du viellissement sur le système cardiovasculaire, en exploitant des données *in vivo*. L'application qui a motivé ce travail est liée à la pharmacologie de sécurité, qui vise à établir, dans une phase pre-clinique, si un médicament est potentiellement dangereux pour la santé [1]. En particulier, nous allons comparer des approches statistiques, d'apprentissage statistique et de modélisation mathématique.

1. Introduction

1.1. Motivation

This work has been motivated by some questions arising in safety pharmacology. Safety pharmacology studies are designed to identify and assess the potential clinical risk of undesirable drug properties before they enter clinical trials, as described in [2].

Many drug development processes must proceed through several stages to be sure for a product to be safe, efficacious, and has passed all regulatory requirements. The preclinical stage encompasses the use of *in vitro* and *in vivo* studies to develop a drug that can safely and effectively be administered for clinical trials. *In vivo*

[1] General Pharmacology Group, Department of Drug Discovery Support, Boehringer Ingelheim Pharma GmbH & Co KG, Biberach an der Riss, Germany
[2] Sorbonne Université and COMMEDIA team, Inria, Paris, France
[3] NOTOCORD, an Instem company, Le Pecq, France
[4] Laboratory of Physiopharmacology, University of Antwerp, Antwerp, Belgium

studies performed in animals are essential to drug development because they have the ability to evaluate the effects a drug has on a living organism. A particular care is taken in assessing adverse effects and drug-drug interactions that cannot be observed *in vitro* [3].

When an animal participates in an experiment in safety pharmacology studies, one can anticipate that the compound tested might have an effect on the organism. It is therefore essential to know when the animal can participate again in an experiment after a sufficient wash-out period. This is of particular importance in cross-over design studies (for details, the reader can refer to [4]). Before an animal will be used in a study, it has to undergo clinical evaluation as well as physiological tests to monitor the condition its cardiovascular system. When these initial tests are successfully performed the particular animal can be labelled as "healthy" and participate in the experiment. Age is one of the factors that has an impact on the function of the cardiovascular system. The effect of age on cardiovascular function in laboratory dogs can be related to decreased blood flow, blood velocity, arterial compliance and distensibility, as well as increased ventricular systolic and diastolic stiffness as a result of prolonged duration of myocardial contractility (see [5], [6], and [7]). In this study, we are going to test the capabilities of several methods to detect the effect of age using some cardiovascular data collected from laboratory dogs.

Among all the questions that pharmacologists and physiologists have, we have tried to answer a pair of questions:

(Q_1) Can we assess if an animal is getting old?
(Q_2) Is it possible to identify an animal by its haemodynamic data by a computer algorithm?

Data about the cardiovascular activity of animals are available over several weeks taken by telemetry. For instance, each hour of recording can be a few gigabytes for each dog. Plus, there are variabilities across several weeks of recording. Therefore, it is necessary to use mathematical methods that can automatically analyse large data sets collected from those initial tests since it is not possible to analyse them manually.

This is a preliminary work to test different methods to answer these questions using a large data set from 4 dogs. Then, the methods can be verified and improved in future work.

1.2. **Methods**

Different viewpoints might be used to address these questions. In this work, we have compared statistical, machine learning, and mathematical modelling methods to analyse some *in vivo* cardiovascular data of 4 dogs. One of the main goals of the project is to assess which methods perform better on these kinds of tasks. In particular, we will compare:

(1) Statistical methods: Physiologists and pharmacologists typically use statistical methods. This technique involves extracting a set of features from the signals and analyzing them statistically. As a first step, we will compute the empirical estimators mean, median, and other statistical criteria to determine if there is any significant difference between the cardiovascular function of young and old animals. Two-tailed Wilcoxon (Mann-Whitney) test will be presented to assess the effect of age on the individual features. The final step in the analysis will be the K-Means clustering (which could be interpreted as an unsupervised learning approach).
(2) Machine learning: Given a database of signals and the outcome of the questions, we can build a map to learn a relationship between the data and the outcome. We will use artificial neural networks, which are typical machine learning algorithms. They were used for instance in ECG analysis [8], cardiac arrhythmia prediction [9, 10], and drug safety studies [11, 12].

(3) Mathematical modelling: By exploiting a priori information about the system, we build a set of equations to simulate the phenomenon under investigation. These equations provide a way to link observable quantities to the outcome. We will use a parametric 0-d model to simulate the global circulation (in the spirit of [13–19]). The parameters of the model, once calibrated, will make it possible to investigate the changes in the animals' cardiovascular system with age.

More precisely, we would like to analyse the advantages and disadvantages of each approach in terms of accuracy and computational cost.

1.3. **Structure**

In section 2 we have presented the experimental data set we have used. In section 3 we have described methods for analyzing the statistics. In section 4 we have explained the Multilayer Perceptron (MLP) method that was implemented to detect the ageing of the dogs and distinguish the dogs. In section 5 we have presented the model for the left part of the heart coupled with a model of the global circulation and the obtained results after calibration. Finally, some conslusions are given in section 6.

2. Experimental Data Sets

In this project, we have used cardiovascular data gathered from 4 dogs: two Beagles (Hexe and Happy) and two Labrador-mix mongrel dogs (Simba and Roxy), involved in safety pharmacology studies for several years. The data are acquired by telemetry from awake and non-anaesthetized animals for many hours (for more details, the reader can refer to [20]). This data set includes values of the Electrocardiogram (ECG), Arterial Pressure (AP), and Left Ventricular Pressure (LVP) signals, recorded every 2 milliseconds. For each dog, we had data corresponding to two different periods of the dog's life: the first one when the dog was 6-7 years old ("*Historical data*"), the second one when it was 8-9 years old ("*New data*"). We have used the cardiovascular data that has been recorded when the animal was younger and "healthy" state (we also call it "Historical" data) to compare with the recent recorded data, helping to decide whether or not this particular animal is "healthy". Each data file includes a seven hours continue recording of a placebo cardiovascular safety pharmacology study (pharmacologically inactive substances [21]). Concerning the measurement of the arterial pressure, according to [22], the signal was sampled either by a catheter in the abdominal aorta or in the femoral artery. The LVP is measured by inserting a catheter connected to a fluid-filled transducer into the left ventricle [23]. From each complete cardiac beat (which in dogs at rest is around 0.8 s), we have extracted 9 quantities (those choices of 9 quantities have been decided by the data provider). All these features have been done by Notocord-Hem™ software. Human supervision and some manual corrections were needed to ensure the features are correctly extracted (especially for ECG). From the ECG we compute (see figure 1):

1 QT interval in ms.
2 QRS interval in ms.
3 RR interval in ms.
4 PR interval in ms.

From the LVP signal we extract the following parameters:

5 Left ventricle systolic pressure, LVP systolic, measured in mmHg.
6 Left ventricle diastolic pressure, LVP diastolic, measured in mmHg.
7 Maximum of the left ventricular pressure, LVP dpdt(max), measured in mmHg s^{-1}.

From the AP signal we extract:

8 Systolic AP, measured in mmHg.
9 Diastolic AP, measured in mmHg.

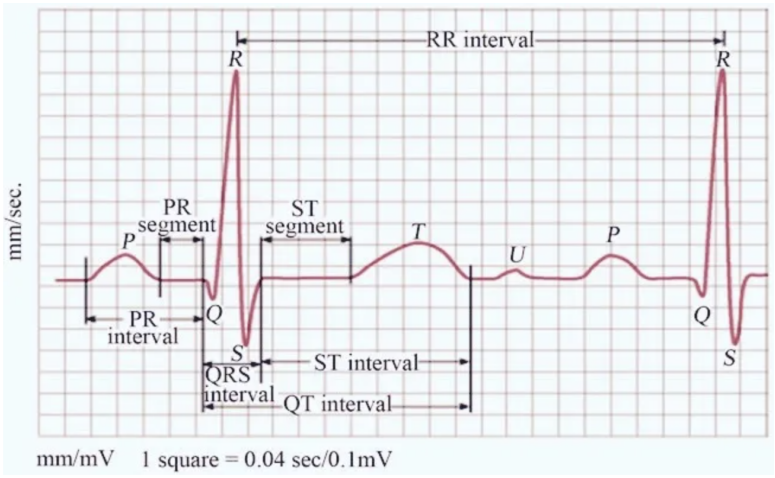


FIGURE 1. Cardiac parameter calculation from the raw ECG signal. See reference [49]

3. Statistical Analysis

In this section, we present the statistical analysis of the data. This has been performed by using solely the 9 features extracted from the telemetry data. In the next section, we will present the methods that have been used and their results.

3.1. **Methodology**

The data set consists of 9 cardiovascular features for every cardiac beat for a total amount of 10887 beats for the young and old animals. We had, henceforth, roughly $N_s = 10^4$ samples. The first analysis consists of computing the mean, first quartile, median, third quartile, and standard deviation for each feature individually by using empirical estimators. Due to the distribution-free nature of the data, a Mann-Whitney U test with a significance level of 0.05 was performed to confirm the results and interpretation provided by the statistical moments estimated (in the spirit of [24]). The goal was to understand whether age is influencing the features, individually. In Mann-Whitney test, the U_1 and U_2 values were computed as

$$U_1 = n_1 n_2 + n_1(n_1 + 1)/2 - R_1,$$
$$U_2 = n_1 n_2 + n_2(n_2 + 1)/2 - R_2.$$

In these formulas, n_1 and n_2 represent the sample sizes for the "Historical" sample and the "New" sample, and R_1 and R_2 represent the sum of the ranks for the "Historical" and the "New" cardiovascular sample, respectively.

As a second method, K-Means clustering has been used in order to determine whether cardiovascular functionality is unique and how age can affect this function (the reader is referred to [25, 26]). This algorithm was used to identify homogeneous subgroups within the data, such that the data points within each cluster are as similar as possible based on euclidean-based distance. Prior to applying the method, we have normalised the feature values and mapped them into the unit hypercube $[0,1]^d$. As a similarity metric, we have used the standard $\ell^{2,d}$ norm (the Euclidean distance in renormalised space). The d value for this study is equal to 9 because we have nine cardiovascular features. The way to assign data points to clusters is to compute the squared distance between them and the cluster centroid (arithmetic mean of all the data points in a cluster) at a minimum.

To perform several tests with different purposes using the K-Means algorithm, we specify a different number of clusters that is estimated in all cases using the clustering gap method, based on [27]. Let $k \in \mathbb{N}$ be the number of clusters, we denote the Euclidean distance between two points D_r, the sum of all pairwise distances would be: $W_k = \sum_{r=1}^{k} \frac{1}{2n_r} D_r$. The gap statistic is defined as:

$$\mathrm{Gap}_n(k) = E_n^*\{\log(W_k)\} - \log(W_k),$$

where E_n^* is the expectation under a sample of size n from the reference distribution. Gap statistic is computed to estimate the most optimized K for clustering. The number of K is selected based on the overall behaviour of uniformly drawn samples, where the greatest jump in within cluster distance occurred. For each number of clusters k, the algorithm compares $\log(W(k))$ with $E_n^*\{\log(W_k)\}$ where the latter is defined via bootstrapping [28]. To eliminate the sampling noise from the data, the optimal K value will only be determined if the change is larger than the others. Figure 3 illustrates how K-Means clustering can estimate that we have two different data sets for the first data union of "Historical" and "New" data of one dog as an example.

As a first step, we have considered, as a data set, the union of the "Historical" and "New" data for each dog. As the second data union, we have shuffled all the cardiovascular data of the dogs in both "Historical" and "New" data. W_k has been used to measure the pertinence of the results derived from K-Means on this database.

In order to determine whether age dominates variability within an individual dog to a greater or lesser extent, we divided the first data union into two and the second data union into four groups.

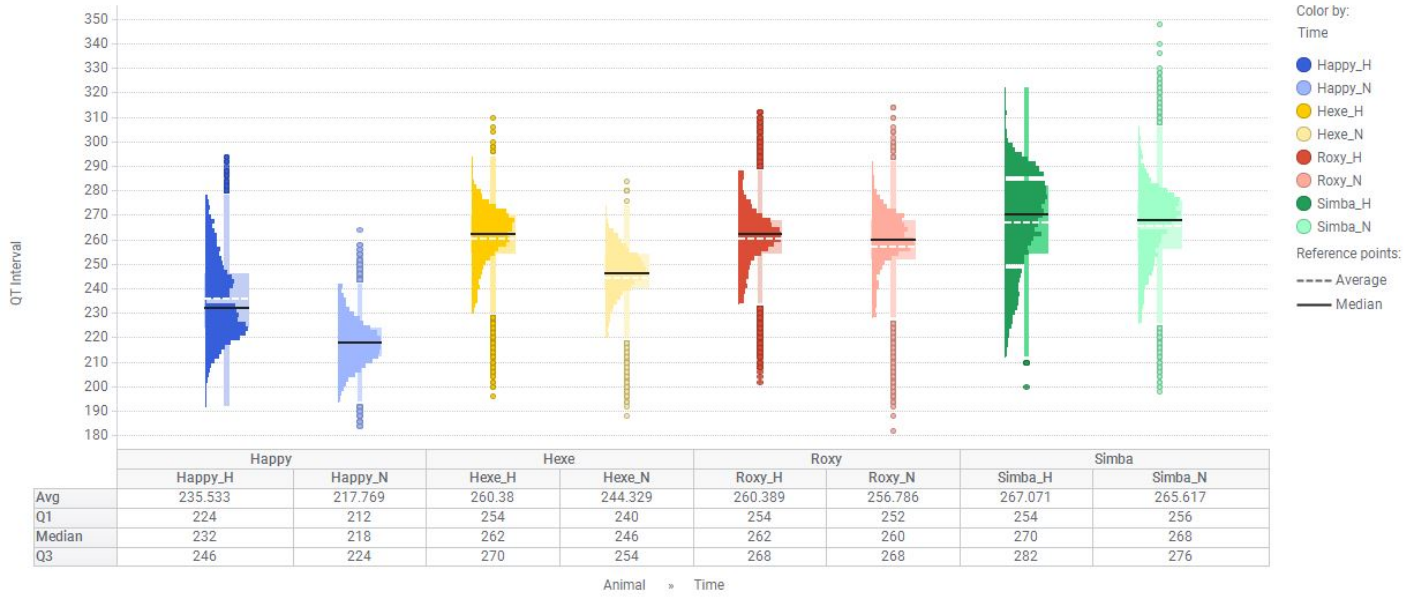
3.2. **Results from Statistical Analysis**

For each cardiovascular feature, we have made box-plot charts to compare the changes over time between "Historical" and "New" data for each dog in order to dynamically analyze the effect of ageing on each cardiovascular characteristic. Figure 2 shows the box-plot charts[1] of the overall QT interval comparison between dogs as an example of the statistical calculation based on ECG features.

3.2.1. *Results answering* Q_1

It has been found that the Mann-Whitney analysis has confirmed that ageing has a significant effect, which has a large and detectable impact on each cardiovascular feature as they relate to ageing.. These results are presented in Table 1.

[1] All the charts have been drawn by TIBCO Spotfire® platform.



	Happy		Hexe		Roxy		Simba	
	Happy_H	Happy_N	Hexe_H	Hexe_N	Roxy_H	Roxy_N	Simba_H	Simba_N
Avg	235.533	217.769	260.38	244.329	260.389	256.786	267.071	265.617
Q1	224	212	254	240	254	252	254	256
Median	232	218	262	246	262	260	270	268
Q3	246	224	270	254	268	268	282	276

FIGURE 2. QT interval overview of all the dogs at two different ages.

TABLE 1. Two-tailed Mann-Whitney U test to determine the effects of ageing on each dog cardiovascular features

Cardiovascular Features	Happy		Simba		Roxy		Hexe	
	U value	p-value	U value	p-value	U value	p-value	U value	p-value
LVP Systolic	$6.19 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$2.83 \cdot 10^7$	$5 \cdot 10^{-3}$	$4.44 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$4.64 \cdot 10^7$	$2.2 \cdot 10^{-16}$
LVP Diastolic	$1.89 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$1.78 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$4.82 \cdot 10^6$	$2.2 \cdot 10^{-16}$	$1.56 \cdot 10^7$	$2.2 \cdot 10^{-16}$
LVP dpdt(max)	$7.47 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$2.95 \cdot 10^7$	$18 \cdot 10^{-15}$	$3.29 \ . \ 10^7$	$1.4 \cdot 10^{-4}$	$1.08 \cdot 10^8$	$1.4 \cdot 10^{-4}$
QT duration	$7.51 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$3.04 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$3.27 \cdot 10^7$	$1.4 \cdot 10^{-3}$	$9.39 \cdot 10^7$	$2.6 \cdot 10^{-2}$
PR duration	$4.42 \cdot 10^7$	$1 \cdot 10^{-4}$	$3.17 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$2.11 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$5.66 \cdot 10^7$	$2.2 \cdot 10^{-16}$
RR duration	$5.65 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$2.82 \cdot 10^7$	$1.1 \cdot 10^{-2}$	$2.39 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$6.29 \cdot 10^7$	$2.2 \cdot 10^{-16}$
QRS duration	$3.39 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$2.53 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$4.97 \cdot 10^6$	$2.2 \cdot 10^{-16}$	$6.78 \cdot 10^7$	$2.2 \cdot 10^{-16}$
APR Systolic	$8.25 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$3.96 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$5.39 \cdot 10^6$	$2.2 \cdot 10^{-16}$	$4.98 \cdot 10^7$	$2.2 \cdot 10^{-16}$
APR Diastolic	$1.93 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$4.01 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$6.02 \cdot 10^7$	$2.2 \cdot 10^{-16}$	$4.36 \cdot 10^7$	$2.2 \cdot 10^{-16}$

To answer Q_1, we have provided Table 2 that shows the impact of ageing on each of our dogs on a case-by-case basis. In order to gain a clearer understanding of the extent to which each dog has been affected by ageing over time, it is necessary to average out the changes between "Historical" and "New" data. It has been observed that Simba's cardiovascular data has remained relatively stable over time. The cardiovascular feature values of Roxy are opposite to the ones of Happy and Hexe (we should also note that Happy and Hexe are Beagles, while Simba and Roxy are Mongrels).

TABLE 2. Total median of "Historical" and "New" cardiovascular data parameter for per each animal and their comparison.

Name	Data	LVP Systolic	LVP Diastolic	LVP dpdt(max)	QT duration	PR duration	RR duration	QRS duration	APR Systolic	APR Diastolic	Avg of all
Happy	Historical	130.73	-4.63	3173.83	232	126	708	40	136.52	90.79	
	New	125.61	-1.83	2746.58	218	126	572	42	123.62	104.47	
	New-Historical	-5.12	2.8	-427.25	-14	0	-136	2	-12.9	13.68	-64.0877
Simba	Historical	114.86	0.48	2075.2	270	114	1256	40	131.99	86.23	
	New	115.72	2.19	2075.2	268	112	1258	40	122.53	76.64	
	New-Historical	0.86	1.71	0	-2	-2	2	0	-9.46	-9.59	-2.0533
Roxy	Historical	98.75	-3.66	1434.33	262	120	1010	46	104.93	81.22	
	New	90.69	1.7	1403.81	260	128	1228	60	92.03	62.83	
	New-Historical	-8.06	5.36	-30.52	-2	8	218	14	-12.9	-18.39	19.2766
Hexe	Historical	117.55	-4.63	3021.24	262	122	978	44	128.62	84.32	
	New	120.48	1.586	2136.23	246	120	832	42	130.21	89.66	
	New-Historical	2.93	6.216	-885.01	-16	-2	-146	-2	1.59	5.34	-114.9926

In the next step, K-Means clustering has been examined on merged "Historical" and "New" data sets for each dog (independently) in order to determine whether or not ageing can comprehensively change the cardiovascular characteristics of individual animals. To avoid over-fitting, data points were grouped into chunks, and for each cardiovascular feature the median of 100 data points was computed.

The positive outcome is the data were roughly clustered according to age in this test. The number of samples that fall into the wrong cluster was computed by assuming "Historical" data are labeled cluster 1 and "New" data are labeled cluster 2. The evaluation of the results of this clustering has been shown in Table 3. This table shows the rates of wrong labeled K-Means clustering for each animal in each cluster of "Historical" and "New". Regarding to this table, the accuracy of K-Means clustering for distinguishing "Historical" and "New" data of Happy, Simba, Roxy, and Hexe are, respectively, 77%, 52%, 87%, and 94%.

TABLE 3. K-Means cardiovascular clustering of one animal in different ages.

	"Historical" wrong labeled / "Historical" cluster	"New" wrong labeled / "New" cluster	Total wrong labeled / Total number
Happy	45/106 = 0.42	0/86 = 0	45/192 = 0.23
Simba	52/84 = 0.61	20/66 = 0.30	72/150 = 0.48
Roxy	7/78 = 0.8	15/82 = 0.18	22/160 = 0.13
Hexe	0/105 = 0	14/103 = 0.13	14/208 = 0.06

3.2.2. *Results answering* Q_2

To check if it is possible to identify an animal by its haemodynamic data and answering Q_2, we have considered a data set containing data from two dogs. The purpose of this test is to examine inter- and intra-dog variability in order to determine if age is more relevant than individual cardiovascular characteristics. The input data was the union of 4 data sets, two dogs in two different ages. As soon as the data set is divided into two clusters by the K-Means algorithm, it is automatically divided by two individual animals' cardiovascular data. This result illustrates that each individual animal has unique cardiovascular characteristics. As shown in Table 4, K-Means clusters the cardiovascular data of each individual dog. The average number of total incorrect classifications was 16 percent, which indicates an 84 percent success rate for distinguishing cardiovascular data between two individual dogs. The K-Means clustering method is thus capable of recognizing differences in age, individual animals, and individual files (each animal in each age group) with acceptable accuracy. Let us point out that if the number of clusters increases, or if the age, gender, and strain of animals are similar, then the accuracy of the K-Means clustering will be reduced.

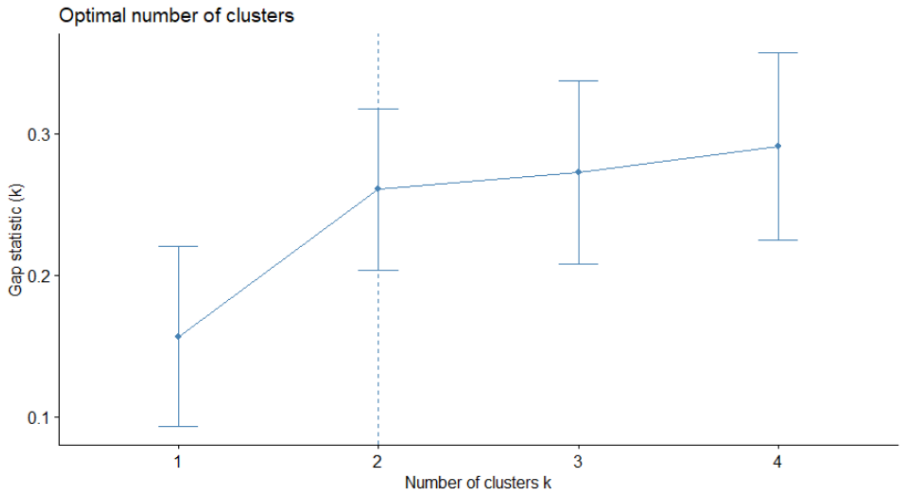


FIGURE 3. K-Means clustering indicates the optimal number of clusters using gap for "Historical" and "New" data combination file for one dog. Gap is within cluster distance, and number of cluster is considering possible cluster number.

TABLE 4. K-Means cardiovascular clustering of two animals in different ages.

Happy-Simba	Happy wrong labeled / Happy cluster	Simba wrong labeled / Simba cluster	Total wrong labeled / Total number
	$22/192 = 0.11$	$20/150 = 0.13$	$42/342 = 0.12$
Happy-Roxy	Happy wrong labeled / Happy cluster	Roxy wrong labeled / Roxy cluster	Total wrong labeled / Total number
	$0/192 = 0$	$16/160 = 0.1$	$16/352 = 0.04$
Happy-Hexe	Happy wrong labeled /Happy cluster	Hexe wrong labeled / Hexe cluster	Total wrong labeled / Total number
	$53/192 = 0.27$	$38/208 = 0.18$	$91/400 = 0.22$
Simba-Roxy	Simba wrong labeled / Simba cluster	Roxy wrong labeled / Roxy cluster	Total wrong labeled / Total number
	$0/150 = 0$	$46/160 = 0.28$	$46/310 = 0.14$
Simba-Hexe	Simba wrong labeled / Simba cluster	Hexe wrong labeled / Hexe cluster	Total wrong labeled / Total number
	$45/150 = 0.3$	$75/208 = 0.36$	$120/358 = 0.33$
Roxy-Hexe	Roxy wrong labeled / Roxy cluster	Hexe wrong labeled / Hexe cluster	Total wrong labeled / Total number
	$46/160 = 0.28$	$0/208 = 0$	$46/368 = 0.12$

4. MACHINE LEARNING ANALYSIS

Considering the answer to Q_1, we would like to detect if the dogs are still healthy by comparing the "Historical" with the "New" data. We can see this problem either as a binary classification or a semi-supervised learning problem. Firstly, we have considered a binary classification. For each dog, the input consists of the set of cardiovascular features introduced in section 2. Given these features, we would like to determine if they were recorded at the younger age (class 0) or at the older age (class 1) for each dog separately. If we get a high classification score, we would conclude that their health conditions have changed. Otherwise, we would consider that they have a similar health condition as when they were young. We have used an MLP to perform this classification.

Secondly, if we consider the ageing assessment as a semi-supervised learning problem, we could use a Replicator Neural Networks (RNN) method on the cardiovascular signals. RNN is usually set up to perform anomaly detection in [29]: the RNN takes the raw signals as input and tries to reconstruct the input itself, as usually done

in autoencoders [30]. In the present context, this would translate as follows: we train an RNN model on the "Historical" signals (normal case). Then we can use trained RNN to reconstruct test samples from "Historical" and "New" signals. By comparing the errors in the reconstructions, we would like to be able to classify whether the signal is coming from a young or old animal. Due to the small changes in the signals and the variabilities between signals, the RNN method is not very successful in performing the detection of dogs' age effects and for sake of brevity, we will not discuss any further the results in this article. However, we have tested the RNN method on another data set reported in the Appendix.

Moreover, we would also like to see if we can distinguish 4 dogs by using their cardiovascular features (answering Q_2). We can view this question as a multi-class classification problem (see [31]). We have tried to classify each dog to see if they are accurately classified in their class[2].

4.1. MLP Method used to Analysis *In vivo* Data

In this part, we have tested MLP to answer Q_1 and Q_2. According to [32], MLP is one of the most commonly used artificial neural networks in medical decision support to help analysing cardiovascular data. MLP consists of multiple layers which include the input layer to receive the features, several hidden layers which are the true computational engines of the MLP, and an output layer that produces a decision or prediction results. MLP is often applied to supervised learning problems especially binary classification. We can train a set of hidden parameters of MLP that are able to learn the relationships between input-output. Then, the trained parameters can be optimized by Back-propagation which computes the gradient of the loss/error with respect to the weights in the network.

We have trained a MLP to discriminate between the "Historical" and "New" state of 4 dogs to answer the first question. "Historical" and "New" data correspond the two classes in the binary classification task. "Historical" and "New" data are mixed in the model training phase. We have used the data set consisting of around 10000 samples of cardiovascular features for each state of each dog to train the MLP model. Those features are normalized using $MinMaxScaler$, explained in [33]. We have used 80% of the samples to train and validate (with cross validation), and 20% of samples to test the MLP model. Our MLP model consists of 3 fully connected hidden layers, which have 9, 6, and, 3 hidden units. The first hidden layer receives 9 cardiovascular features (LVP Systolic, LVP Diastolic, LVP dpdt(max), RR interval, PR duration, QT interval, QRS duration, APR Systolic, APR Diastolic). The outputs from each hidden layer are transformed by a ReLU function and the results in the output layer will be transformed to a decision boundary by a *sigmoid* function.

4.2. Results from MLP Method Answering Q_1

To evaluate the performance of the MLP method, we have computed the success rate which is defined as the number of correctly predicted samples divided by the total number of samples of the dogs as in Table 5. We have high success rates, 98.24%, 87.40%, 97.81%, and 98.33% to discriminate between the "Historical" and "New" state of our 4 dogs. As Simba has the lowest success rate than other dogs, maybe Simba has fewer changes in its cardiovascular performance compared to the other 3 dogs.

TABLE 5. Results from MLP model to discriminate "Historical" and "New" state of the dogs.

	Happy	Simba	Roxy	Hexe
Success Rate	98.24%	87.40%	97.81%	98.33%

[2]All the Machine Learning models were implemented in Python using $TensorFlow^{TM}$.

4.3. **Results from MLP Method Answering Q_2**

We also would like to check if we can identify a dog by using an MLP, to perform a multi-class classification. According to [34], multiclass classification makes the assumption that each sample is assigned to one and only one label. We have labelled 4 classes Happy, Simba, Roxy, and Hexe. If there are significant differences in their cardiovascular features for each dog, we expect the model will have a high score to label them correctly, which means we can identify a specific dog among the 4. In the multi-class classification, we have used the same cardiovascular features as input. The difference between MLP used for binary classification and multi-class classification is in the output layer, the number of outputs being equal to the number of classes and the results will be transformed by a *softmax* function. In the training and testing phase, only "Historical" data of 4 dogs were used to train and test the multi-class classification model. From the results reported in figure 4, we can conclude that over 93% of the time, we can identify one dog among 4 dogs and 7% of the time, we miss identifying the dog.

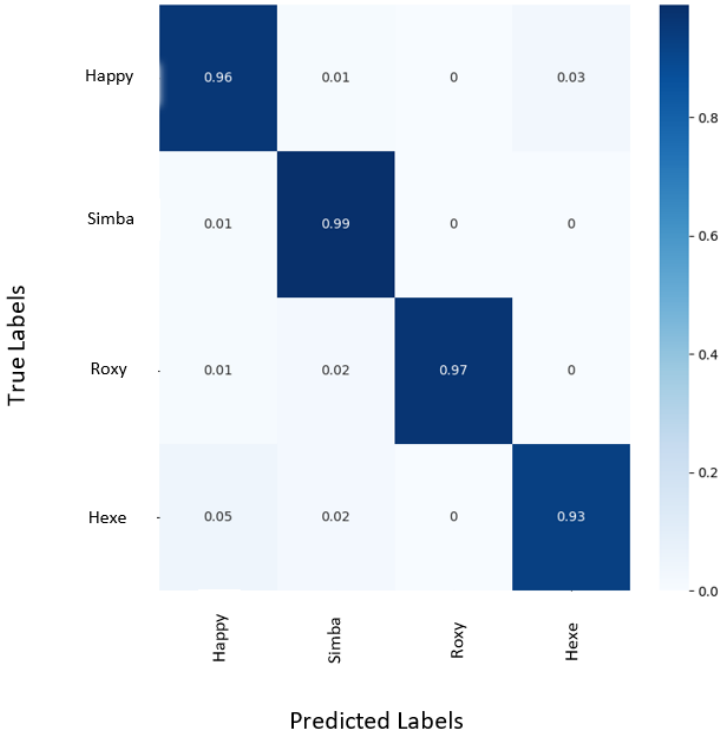


FIGURE 4. Classification results for identifying dogs using MLP method.

5. MATHEMATICAL MODELLING ANALYSIS

In this part of the project, we have considered a 0-d model (concerning 0-d models the reader is referred to [13–19]) for the left part of the heart coupled to systemic circulation which makes it possible to describe the main observable haemodynamics quantities and their evolution in time. Once the data are reproduced, we can calibrate the model by estimating the values of the parameters. The calibration is performed for the "Historical" state and for the "New" state of the animal and the values for the parameters in each situation will help to assess if the dog has changed significantly with age.

5.1. **Analog Circuit Model for the Left Ventricle**

The heart beat (concerning the physiology the reader is referred to [35]) is a two stage pumping action over a period of about one second subdivided in 2 stages: systole and diastole. During systole, the pressure in the left ventricle increases, exceeding the left atrium pressure. Then the mitral valve closes and the aortic valve opens and the blood flows into the aorta and out to the rest of the body. In diastole, the rate of contraction of the myocardium begins to slow and the aortic valve closes. When the ventricle relaxes, the pressure in the left ventricle falls and when it decreases below the pressure in the left atrium, the mitral valve opens and this lets blood flow from the left atrium to the ventricle.

The goal is to build a 0-d model of the left part of the heart coupled to system circulation by an electrical analogy [36].

The $0-d$ electric circuit model (see [19]) for the left ventricle is shown in figure 5.

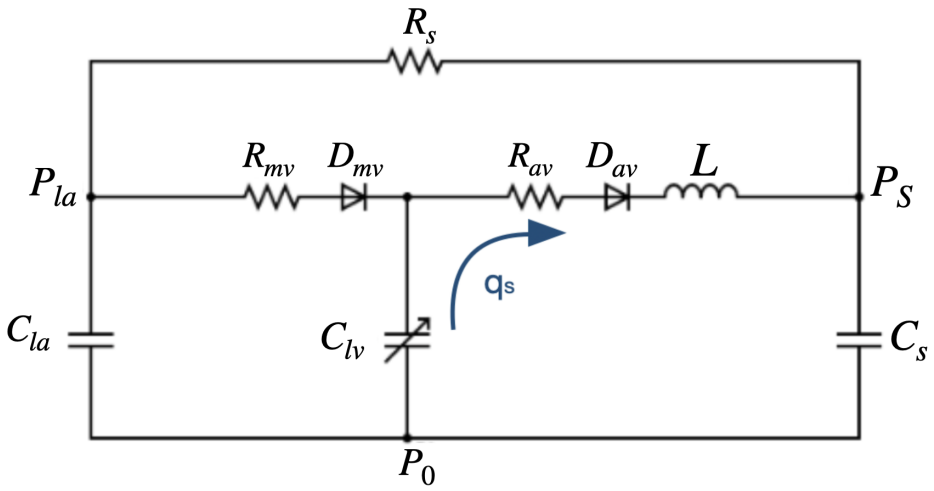


FIGURE 5. Electric circuit model.

The valves are represented by diodes in series with a resistor: R_{mv} and D_{mv} for the mitral valve and R_{av} and D_{av} for the aortic valve and R_s is the resistance of systemic circulation. The coil L represents the inertia of blood in the aorta. The compliances of the left atrium and systemic circulation are represented by C_{la} and C_s, respectively. We have used a time-varying left ventricular compliance $C_{lv}(t)$ to represent the action of the heart muscle. The elastance $E(t)$ is the reciprocal of the compliance ($E(t) = 1/C_{lv}(t)$) and it represents the contractile state of the left ventricle. It relates to the ventricle's pressure and volume (detailed in [37]) according to the expression:

$$E(t) = \frac{P_{lv}(t)}{V_{lv}(t) - V_0},$$

where $P_{lv}(t)$ is the left ventricular pressure, $V_{lv}(t)$ is the left ventricular volume and V_0 is a reference volume. The time variations in this model are due to the cyclical nature of the ventricle elastance and it changes as a function of time within one cardiac cycle. In our case, we are using the ECG to model the activation of the ventricle and, more precisely, the peaks of the QRS complex. The QRS complex (see figure 1) corresponds to the electrical forces generated by ventricular depolarization and represents the pumping action of the ventricles.

From the ECG, we save the times at which the R peaks of the QRS complex occur and this is what we call the *activation times* of the ECG.

We approximated $E(t)$ by the following expression:

$$E(t) = \sum_{k=1}^{m} a \cdot [\tanh[b \cdot (t - w_k)] - \tanh[b \cdot (t - w_k - d)]] + h,$$

where $m \in \mathbb{N}^*$, a, b, d and h are constants, $t \in \mathbb{R}$ is the time and $w \in \mathbb{R}^m$ are the activation times of the ECG.

In order to get the equations for the circuit, three different cases have been considered:

- **Case 1: Filling.** The mitral valve is **opened** and the aortic valve is **closed** so the left ventricle is being filled.
- **Case 2: Ejection.** The mitral valve is **closed** and the aortic valve is **opened** so blood is being ejected from the left ventricle.
- **Case 3: Isovolumic phase.** Both valves are **closed** and the capacitor has either just being filled (isovolumic contraction) or emptied (isovolumic relaxation).

Each phase of operation of the left ventricle can be modeled by a system of linear ordinary differential equations (ODEs). However, the whole system is non-linear because each diode has two states so we have three different cases (first diode is on and second one is off, first diode is off and second one is on, or both diodes are off). Each case is modeled by a different equivalent circuit since when a diode is off, it acts as a wire and we have open-circuit and when it is on, we have short-circuit.

The system of ODEs has been derived by applying Kirchoff's Current Law and Ohm's Law (the reader is referred to [38] for an overview on the topic) to the analog circuit model as follows:

CASE 1: If $P_{la}(t) > P_{lv}(t)$ and $P_{lv}(t) < P_s(t)$

For the first node, with left atrial pressure (LAP) $P_{la}(t)$, we have that

$$C_{la}\frac{dP_{la}}{dt} = \frac{P_{lv} - P_{la}}{R_{mv}} - \frac{P_{la}}{R_s}.$$

For the one in the middle with value $P_{lv}(t)$

$$C_{lv}\frac{dP_{lv}}{dt} + P_{lv}\frac{dC_{lv}}{dt} - P_0\frac{dC_{lv}}{dt} = \frac{P_{la} - P_{lv}}{R_{mv}},$$

and since $E_{lv}(t) = 1/C_{lv}(t)$, we can rewrite the previous expression as

$$\frac{dP_{lv}}{dt} = E\left(\frac{P_{la} - P_{lv}}{R_{mv}}\right) + \frac{P_{lv}}{E}\frac{dE}{dt} - \frac{P_0}{E}\frac{dE}{dt},$$

where $P_0 \in \mathbb{R}$ is just a reference value. Then, for the node on the right, with arterial pressure $P_s(t)$,

$$C_s\frac{dP_s}{dt} = \frac{P_{la} - P_s}{R_s},$$

and, for the aortic flow $q_s(t)$

$$\frac{dq_s}{dt} = 0.$$

The system of equations for the filling phase is then

$$\begin{cases} C_{la}\dfrac{dP_{la}}{dt} &= \dfrac{P_{lv}-P_{la}}{R_{mv}} - \dfrac{P_{la}}{R_s}, \\ \dfrac{dP_{lv}}{dt} &= E\left(\dfrac{P_{la}-P_{lv}}{R_{mv}}\right) + \dfrac{P_{lv}}{E}\dfrac{dE}{dt} - \dfrac{P_0}{E}\dfrac{dE}{dt}, \\ C_s\dfrac{dP_s}{dt} &= \dfrac{P_{la}-P_s}{R_s}, \\ \dfrac{dq_s}{dt} &= 0. \end{cases}$$

CASE 2: If $P_{lv}(t) > P_{la}(t)$ and $P_{lv}(t) > P_s(t)$

Proceeding in an analogue way, we get the following system of ODEs

$$\begin{cases} C_{la}\dfrac{dP_{la}}{dt} &= \dfrac{P_s-P_{la}}{R_s}, \\ \dfrac{dP_{lv}}{dt} &= \dfrac{P_{lv}}{E}\dfrac{dE}{dt} - Eq_s - \dfrac{P_0}{E}\dfrac{dE}{dt}, \\ C_s\dfrac{dP_s}{dt} &= \dfrac{P_{la}-P_s}{R_s} + q_s, \\ \dfrac{dq_s}{dt} &= \dfrac{P_{lv}-P_s}{L} - \dfrac{q_s}{L}R_{av}. \end{cases}$$

CASE 3: If $P_{lv}(t) > P_{la}(t)$ and $P_{lv}(t) < P_s(t)$

In this case, both valves are closed (both diodes are off) and there is no current flow. Then the equations are

$$\begin{cases} C_{la}\dfrac{dP_{la}}{dt} &= \dfrac{P_s-P_{la}}{R_s}, \\ \dfrac{dP_{lv}}{dt} &= \dfrac{P_{lv}}{E}\dfrac{dE}{dt} - \dfrac{P_0}{E}\dfrac{dE}{dt}, \\ C_s\dfrac{dP_s}{dt} &= \dfrac{P_{la}-P_s}{R_s}, \\ \dfrac{dq_s}{dt} &= 0. \end{cases}$$

5.2. **Results from Mathematical Modelling**

To solve the equations in each case according to time we have used a backward differentiation formula (BDF) solver that is implemented in the Python built-in solver *odeint* (see [39] and [40] for more details). In order to solve the model, we need to give as **input** the parameters $(R_s, R_{mv}, R_{av}, C_a, C_s, L, P_0, a, h)$ and the peaks of the ECG. Then we get as **output** the LVP, the AP, the LAP and the aortic flow.

To obtain the optimized parameters for each dog we have used the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) optimization algorithm that is implemented in Python in the *pycma* module [41]. We need to give as **input** an initial estimation of the parameters and the experimental data for the AP, LVP and ECG. The initial standard deviation was chosen as $\sigma_0 = 0.4$ and we have used a population size 1000 times larger than the default value. The objective function needs also to be specified. We have chosen the following function to be minimized

$$J = \sqrt{\sum_{i=1}^{n} \left| P_{lv}^{(i)} - \hat{P_{lv}}^{(i)} \right|^2 + \left| P_s^{(i)} - \hat{P_s}^{(i)} \right|^2},$$

where $P_{lv}^{(i)}$ is the experimental value of the left ventricle pressure at each time, $\hat{P_{lv}}^{(i)}$ is the predicted value of the left ventricle pressure at each time, $P_s^{(i)}$ is the experimental value of the arterial pressure at each time, $\hat{P_s}^{(i)}$ is the predicted value of the arterial pressure at each time and $n \in \mathbb{N}^*$.

The **output** of the CMA-ES algorithm are the optimized parameters of the model.

5.2.1. *Results answering Q_1*

For each dog, we have estimated the parameters for the "Historical" and "New" state. Table 6 shows the optimized model parameters for each dog at both times. The solution of the model for its corresponding parameters for each dog is plotted in figures 6, 7, 8 and 9.

We have computed the difference between the values predicted by our model and the values observed. For that purpose, we have found the local maxima of the LVP and we have computed the difference between the real signal and the simulated one at each peak (local maxima) for the LVP. The relative error is

$$\xi(P_{lv}) = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i)^2}},$$

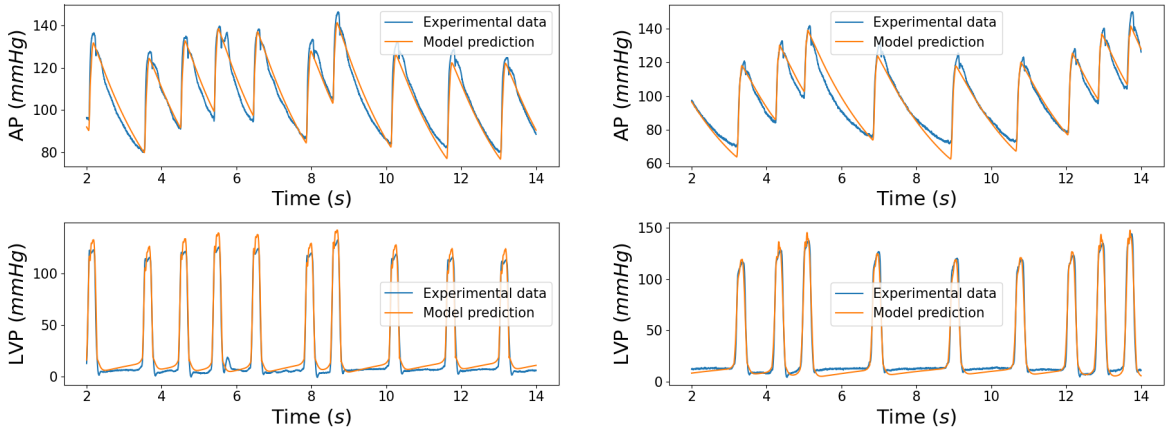
where y is the value of the pressure at each peak in the true signal, $\hat{y}$ is the value in the predicted signal and $n \in \mathbb{N}^*$. For the AP we have taken into account not only the difference between maxima, but also the one between minima

$$\xi(P_s) = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i)^2}} + \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(w_i - \hat{w_i})^2}{\frac{1}{n}\sum_{i=1}^{n}(w_i)^2}},$$

where w is the value of the pressure at each minimum peak in the true signal, $\hat{w}$ is the value of the pressure at each minimum peak in the predicted signal and $n \in \mathbb{N}^*$.

TABLE 6. Model parameters (from left to right) for Happy "Historical" (2015), Happy "New" (2018) , Simba "Historical" (2018), Simba "New" (2021), Roxy "Historical" (2018), Roxy "New" (2021), Hexe "Historical" (2018) and Hexe "New" (2020). The relative error for the LVP and for the AP is also shown for each dog in both states ("Historical" and "New").

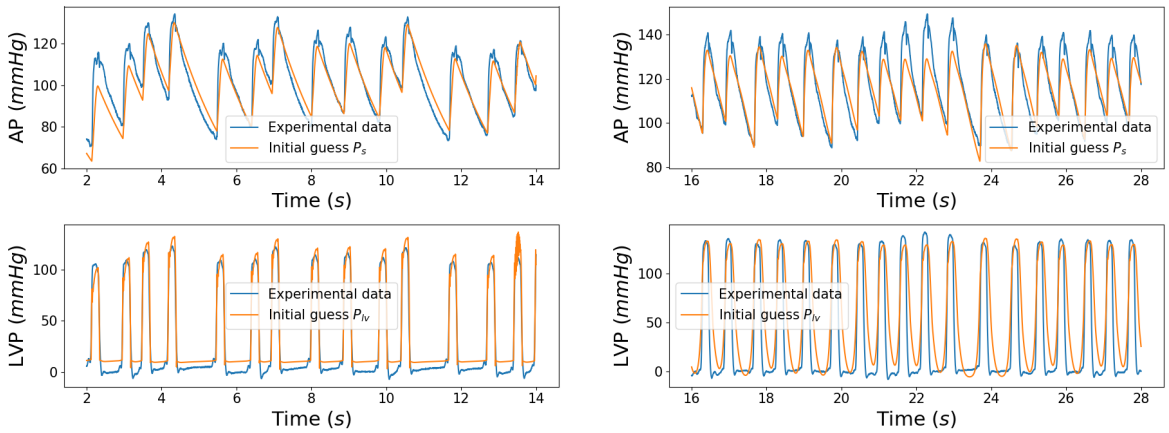
Parameters	**Happy "H"**	**Happy "N"**	**Simba "H"**	**Simba "N"**	**Roxy "H"**	**Roxy "N"**	**Hexe "H"**	**Hexe "N"**
R_s [mmHg $\cdot$ s/cm^3]	$1.62 \cdot 10^{-3}$	$1.20 \cdot 10^{-1}$	$8.06 \cdot 10^{-2}$	$2.67 \cdot 10^{-1}$	8.78	3.87	$1.45 \cdot 10^{-3}$	$6.26 \cdot 10^{-2}$
R_{mv} [mmHg $\cdot$ s/cm^3]	$2.06 \cdot 10^{-6}$	$1.75 \cdot 10^{-2}$	$5.12 \cdot 10^{-5}$	$1.47 \cdot 10^{-4}$	$1.26 \cdot 10^{-1}$	$1.12 \cdot 10^{-2}$	$7.02 \cdot 10^{-7}$	$7.38 \cdot 10^{-3}$
R_{av} [mmHg $\cdot$ s/cm^3]	$1.85 \cdot 10^{-5}$	$1.18 \cdot 10^{-3}$	$6.33 \cdot 10^{-4}$	$2.45 \cdot 10^{-3}$	$3.57 \cdot 10^{-4}$	$8.18 \cdot 10^{-4}$	$3.55 \cdot 10^{-6}$	$2.95 \cdot 10^{-4}$
C_a [cm^3/mmHg]	$3.65 \cdot 10^{4}$	7.01	$1.36 \cdot 10^{2}$	$5.18 \cdot 10^{1}$	$2.61 \cdot 10^{-1}$	1.72	$4.59 \cdot 10^{4}$	$1.0808 \cdot 10^{1}$
C_s [cm^3/mmHg]	$1.21 \cdot 10^{3}$	9.39	$2.99 \cdot 10^{1}$	9.27	$3.26 \cdot 10^{-1}$	$9.67 \cdot 10^{-1}$	$1.71 \cdot 10^{3}$	$1.63 \cdot 10^{1}$
L [mmHg $\cdot$ s^2/cm^3]	$1.68 \cdot 10^{-7}$	$9.18 \cdot 10^{-6}$	$1.1151 \cdot 10^{-5}$	$5.26 \cdot 10^{-5}$	$1.98 \cdot 10^{-4}$	$1.60 \cdot 10^{-4}$	$6.30 \cdot 10^{9}$	$3.41 \cdot 10^{-6}$
P_0 [mmHg]	$1.66 \cdot 10^{4}$	$6.41 \cdot 10^{1}$	$3.05 \cdot 10^{3}$	$7.71 \cdot 10^{2}$	$1.63 \cdot 10^{-1}$	$9.52 \cdot 10^{-1}$	$2.25 \cdot 10^{3}$	$3.16 \cdot 10^{1}$
a [mmHg/cm^3]	$9.08 \cdot 10^{-3}$	$6.90 \cdot 10^{-1}$	$9.27 \cdot 10^{-2}$	$1.95 \cdot 10^{-1}$	$1.07 \cdot 10^{1}$	2.39	$5.36 \cdot 10^{-3}$	$2.87 \cdot 10^{-1}$
h [mmHg/cm^3]	$1.12 \cdot 10^{-2}$	0.00	$6.92 \cdot 10^{-2}$	$6.07 \cdot 10^{-2}$	$3.36 \cdot 10^{-3}$	$3.87 \cdot 10^{-3}$	$1.23 \cdot 10^{-3}$	$1.40 \cdot 10^{-6}$
Relative errors								
$\xi(P_s)$	10.83 %	10.56 %	6.39 %	9.39 %	5.22 %	8.53 %	10.13 %	9.51 %
$\xi(P_{lv})$	8.38 %	4.11 %	8.94 %	3.99 %	3.51 %	7.68 %	6.17 %	4.17 %



(a) AP and LVP for Simba 2018 ("Historical")

(b) AP and LVP for Simba 2021 ("New")

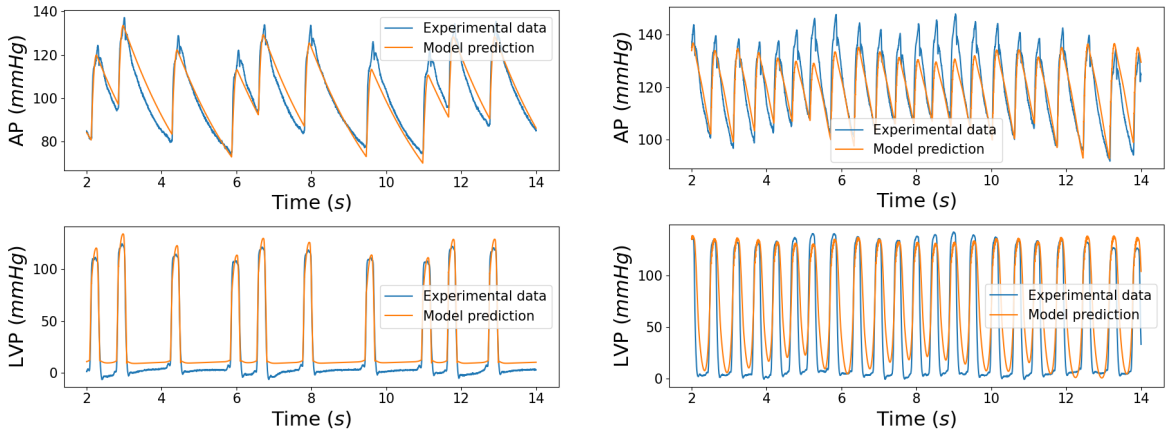
FIGURE 6. Experimental data compared with the model prediction for Simba.



(a) AP and LVP for Happy 2015 ("Historical")

(b) AP and LVP for Happy 2018 ("New")

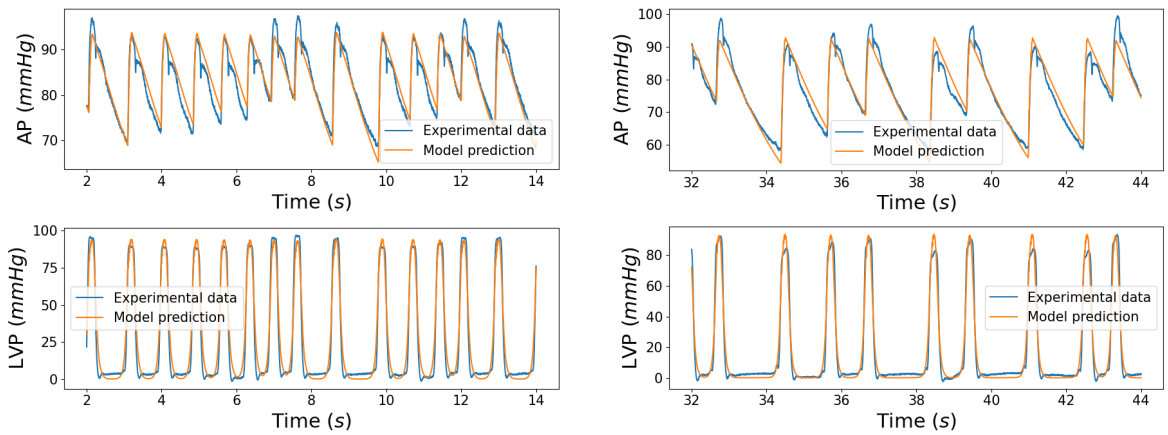
FIGURE 7. Experimental data compared with the model prediction for Happy.



(a) AP and LVP for Hexe 2018 ("Historical")

(b) AP and LVP for Hexe 2020 ("New")

FIGURE 8. Experimental data compared with the model prediction for Hexe.



(a) AP and LVP for Roxy 2018 ("Historical")

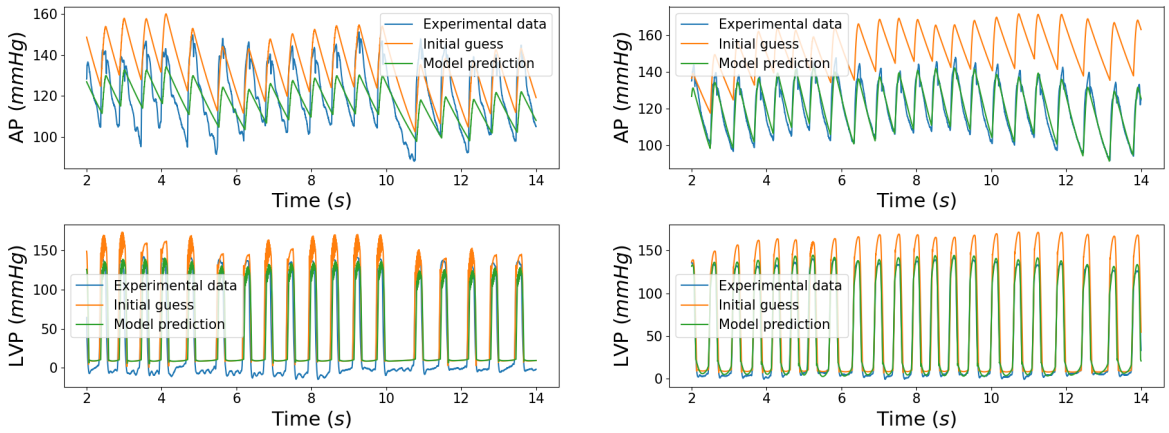
(b) AP and LVP for Roxy 2021 ("New")

FIGURE 9. Experimental data compared with the model prediction for Roxy.

We want to address if the animals are getting old or not by the change in the parameters in the electrical model. To check that, we have made the following tests:

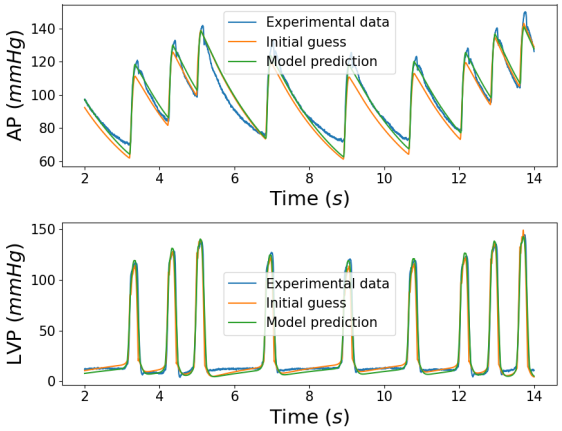
(1) We have computed the optimized parameters of the model for each dog in the "Historical" state and also in the "New" state (they are shown in Table 6).
(2) Afterwards, we have solved the 0-d model for each dog in the "New" state considering as initial guess the optimized parameters for the "Historical" state.
(3) We have also directly solved the model in the "New" state, with the parameters of the "Historical" state.

The purpose is to see if we can match the real data in the "New" state with the optimized parameters from the "Historical" state. Some results are shown in figure 10.



(a) AP and LVP for Happy 2018 starting with parameters of Happy 2015

(b) AP and LVP for Hexe 2020 starting with parameters of Hexe 2018



(c) AP and LVP for Simba 2021 starting with parameters of Simba 2018

FIGURE 10. Experimental data compared with the model prediction when we start the optimization from the optimized "Historical" parameters as initial guess.

After repeating the simulations for several chunks of beats, we have observed that Simba barely changes. With the optimized parameters in the "Historical" state we can reproduce the data in the "New" state. However, we have observed significant changes for Happy, Roxy and Hexe. In Happy and Hexe we have observed that as the dogs are getting older the resistances increase and the capacitances decrease as a general tendency. The capacitance gives us the ability of the vessels to get elastic and since the capacitances decrease, the vessels become less elastic with the years. On the contrary, we have observed in Roxy a different tendency. Roxy has experienced a decrease in the resistances and an increase in the capacitances which is the opposite behaviour as in Hexe and Happy and may be interpreted as her features have improved with age.

To answer properly Q_1, we need to set a threshold of the error that determines that the animal is changing with age. In other words, at which relative error do we claim that the "Historical" parameters do not fit the

"New" state because the dog has changed? To establish that threshold, we have solved the model for each dog for several chunks with the parameters of Table 6. The maximum of the error that we got for each animal, is going to be our threshold to claim whether the dog experienced changes with age or not. The thresholds to determine if the model is fitting the data are 14 % (Happy), 15 % (Simba), 16 % (Roxy) and 13 % (Hexe).

To study the accuracy of the mathematical modelling method answering Q_1, we have solved the model for each dog with our calibrated parameters in other chunks of beats than the ones used for setting the threshold. First, we have considered chunks in the "Historical" state and we have used the "Historical" parameters to solve the model (same applies for the "New" state). If we get a relative error that is less than the threshold, then it means that the model has worked. But if we get a relative error that is bigger than the threshold, it means that the model suggests that the dog changed but it is a mistake since we are just using the calibrated "Historical" parameters but for a different chunk. We have repeated this computation for a certain number of chunks and we have computed the success rate (number of correctly predicted chunks divided by the total number of tested chunks), that was always higher than 95 %.

Therefore, we can conclude that we are able to identify from the parameters if a dog is changing with age or not.

5.2.2. *Results answering Q_2*

The second question that we wanted to answer is if we can discriminate between dogs. If we look at the optimized parameters of the model that are shown in Table 6, we can see that they substantially change between different dogs. We have solved the model at a given moment with the data of a given dog but using the optimized parameters for the same moment but for another dog. The result of doing that for every dog, was that the relative error in the left ventricular pressure ($\xi(P_{lv})$) was sometimes below the values of the thresholds but the value for the relative error in the arterial pressure ($\xi(P_s)$) was always above 32 % (higher than the thresholds).

Although it would be necessary to do the study with a larger number of chunks in order to withdraw a strong conclusion, it is already possible to spot some differences between the dogs since the parameters of one dog never fit the data of another dog.

6. Conclusion and Discussion

In this paper, we have tried to analyse some *in vivo* experiments data for addressing the effect of ageing in an animal by using statistical, machine learning and mathematical modelling methods. On the whole, we are able to identify some changes in dogs' cardiovascular data in the "New" state compared to the "Historical" state. Moreover, we can discriminate not only between "New" and "Historical" state of the dog, but also distinguish one dog among the others.

In terms of accuracy, cost, and robustness, we have compared the advantages and disadvantages of each method. Although the way how we answer the questions is very different in mathematical modelling and in statistical and machine learning approaches, we can set a link on how to compare the accuracy of the methods. With respect to the answer to question Q_1 in the statistical method, changes between the "New" and "Historical" state of cardiovascular data for each dog in a positive or negative direction are always measurable, as shown in the statistical results. However, the K-Means clustering method could recognize and cluster these changes for Happy, Simba, Roxy, and Hexe with an accuracy of 77%, 52%, 87%, and 94%, respectively. Machine learning method can correctly (above 97% for Happy, Roxy, and Hexe) distinguish between the "New" and "Historical" state of the dog. When "New" and "Historical" states are more similar for Simba, the error of the machine learning method would be higher (12.60%). Regarding mathematical modelling, we can see that the relative

error between the values predicted by our model with the optimized parameters and the values observed, is typically below the thresholds. The accuracy of this approach was above 95 %.

According to question Q_2, K-Means clustering accurately identified 84% of dogs among four cardiovascular data of two dogs of different ages. Machine learning method can identify one dog among 4 dogs with more than 93% accuracy. To test the accuracy of the identifiability of the dogs with the mathematical modelling approach, we had the same logic as the one with ageing but we have tried the calibrated parameters of one dog on another dog and see if they could fit. It was never the case so the accuracy in this case is around 100 %.

Regarding the cost and robustness, K-Means and MLP use the extracted cardiovascular features as input and they would require computational cost in the feature extraction process. Considering that the number of samples and the size of the input is not large, K-Means only takes very small computational cost. We could also train a MLP with high accuracy (above 93%) and very little training cost (less than 20 minutes). In particular, MLP is robust to beats to beats variability and provides a promising result for the above biological questions. Regarding the computational cost for mathematical modelling, it requires a light pre-processing of the data. Before running the optimization, the only thing we need to do is to extract the activation times of the ECG and this is almost instantaneous. However, the optimization process to obtain the parameters of the model takes on average 7 hours, although it depends a lot on the initial guess that we consider. Once we have the parameters, model integration in time is cheap from a computational point of view (few seconds). Regarding the robustness of the parameter estimation procedure in mathematical modelling, we have tried so far to run the model with the same parameters for other intervals to see if we are able to reproduce other windows of time with the parameters fitted from one chunk. Usually, we are able to fit other time intervals if we do not go very far from the original one. It is also important to take into account that if the activity of the dog changes (if it is playing, eating, making digestion...) the parameters could change. In the future, we would like to get the model parameters for a larger number of chunks and analyze their variability of them in the “New” and "Historical" states in order to give a reliable interval for the parameters in each state and to have a better notion of the accuracy of this approach. Same applies for the identifiability of each dog in order to know what is the interval in which the parameters move in each dog.

6.1. **Limitations and Perspectives**

The limitation of the statistical algorithms is that when the number of animals is high or strains and genders of dogs are similar, the accuracy may decrease. Statistical methods are modelled using extracted features which require computational resources and the number and choices of features may have an impact on the analysis of the results (missing some information from the raw signals).

Similarly, MLP needs extracted features as input which other types of neural networks like Convolutional Neural Network (CNN) in [42] do not. CNN can process and learn the important parameters from the signals directly to perform a binary classification. By using CNN, we can avoid the feature extraction process step. However, as the dimension of the signal data is usually high, the CNN method would need more training costs. We can also consider the autoencoder method calibrated by some cardiovascular features to do anomaly detection like in [44] to perform a semi-supervised learning. In general, the neural network method has a "black box" nature. This method can’t help us to understand the mechanism of dogs’ cardiovascular system. So we won’t be able to use it to answer some questions like how much and which element of cardiovascular function has changed. However, in a mathematical modelling approach, the parameters have a link with age which can give us some information related to how age changes cardiovascular functionality (that is to say if the dog gets older or younger). Moreover, if the number of dogs increases in the future, which means the number of classes also increases, the accuracy of correctly identifying the dogs decreases for machine learning and statistical methods. To test the neural network method and statistical method for a larger number of dogs, we need to find a training strategy that can include enough dogs and acceptable model accuracy. For the mathematical modelling method,

it is necessary to do the optimization of the parameters for each dog in each state ("Historical" and "New") and this process takes a lot of time. Consequently, if the number of dogs is high, it will be computationally very demanding in terms of time to answer the fore-mentioned questions.

In the future perspective, we would like to take into account the data of more dogs to test our methods. It is also very important from the mathematical modelling point of view to do the study with more chunks to be able to perform classification tasks on the parameters. We would also like to work on more experiments with different species and strains to gather more information about the effect of ageing on cardiovascular performance. Discovering the impact factors on cardiovascular functionality helps us to understand the reason why some animals have more changes in their cardiovascular data by comparing with others.

Acknowledgements

Elham, Sara and Haibo are ESR-fellows of the INSPIRE European Training Network. INSPIRE receives funding from the EU Horizon 2020 Research and Innovation programme, under the Marie Skłodowska-Curie GA 858070 [1].

We appreciate all support from Pieter-Jan Guns as the scientific coordinator of INSPIRE European Training Network, Michael Markert, Georg Rast, Karin Graf, Jessica Schiwon, Thomas Trautmann and Florian Krause from Boehringer Ingelheim Pharma GmbH & Co KG, Damiano Lombardi from Inria, Muriel Boulakia from Université Paris-Saclay, UVSQ, CNRS, Laboratoire de Mathématiques de Versailles and Sylvain Bernasconi from NOTOCORD and Instem companies.

Appendix

A. RNN Method Used for Anomaly Detection

We have a second experimental data set concerning the field potentials (FP) recordings of hiPSC-CMs obtained by multi-electrodes arrays (MEA) technology. Specifically, the experimental data set corresponds to FP signals recorded using 96 wells MEA plates, each well containing 8 recording electrodes and 12 drugs have been tested in the MEA plates. Each drug has been added at four concentrations in four different wells and this process is replicated five times. Among all the collected FP signals, we observed that some signals may have an altered shape compared to the majority of the signals because of poor cells and recording conditions. Those poor recordings can bring biases to the conclusion of the experiments and hence they need to be removed from the data set. This is usually done by manually selecting those poor recordings usually which takes a lot of resources. In this work, we try to investigate methods that can automatically and reliably detect the poorly recorded signals in MEA data set. We have tried RNN method to answer this question.

RNN is an example of autoencoder. Since autoencoders were first proposed in [29], they have been commonly used in outlier detection and anomaly detection [43]. According to [44], autoencoder is a kind of feed-forward neural network with a systemic structure that is composed of multiple hidden layers which include encoding and decoding parts. The input and output layers have the same size (figure 11). The encoder compresses the input data to some latent representations. The decoder decodes the compressed latent representations and reconstructs the input data. The autoencoder can consist of multiple hidden layers. The output from each hidden layer will be transformed by an activation function. The parameters of the RNN are optimized by

minimizing the reconstruction errors. The error is measured as Relevant Mean Squared Error:

$$R - MSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{m}(y_i)^2}}$$

Where $m \in \mathbb{N}^*$ and $y, \hat{y} \in \mathbb{R}^m$ are the true and reconstructed signals respectively. Since the difference of amplitudes for normal and abnormal signals are large, if we only consider Absolute Mean Squared Error, we are not able to measure the fair differences between true and reconstructed signals. Henceforth we considered the relative $\ell^{2,m}$ errors between the signal and its reconstructions. In this way, the value of errors will consider the scale of the signals.

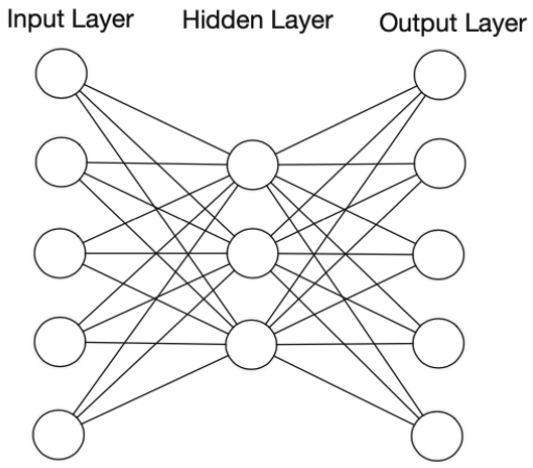
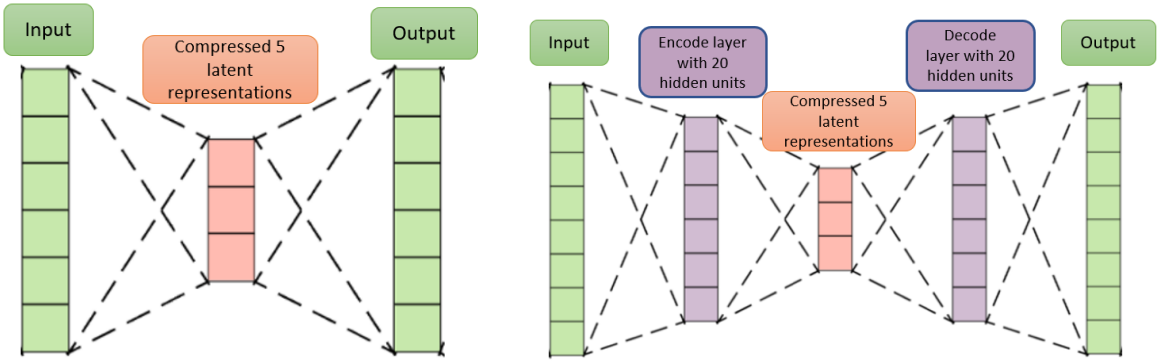


FIGURE 11. Autoencoder systemic structure.

In this work, we have used 34531 signals to train (80% of the signals) and validate (20% of the signals) an autoencoder to reconstruct the input signals. Then the autoencoder has been used to reproduce the unseen examples (4364 abnormal signals have not been used in the training set). By comparing the errors of the reconstructed signals between the normal and abnormal signals, we expect the autoencoder to be able to detect the abnormality. We have tested 2 architectures of the autoencoder shown in figure 12.

A.1. **Results from Autoencoder for Anomaly Detection**

We have listed the results from two architectures for autoencoder in Table 7. The second architecture could reduce the errors of the reproduced signals for the training set by 0.019 compared to the first architecture. Both architectures could reconstruct the normal signal with lower errors. In contrast, both architectures have more difficulties to reproduce the abnormal signals in the test set as shown in figure 13 and figure 14.



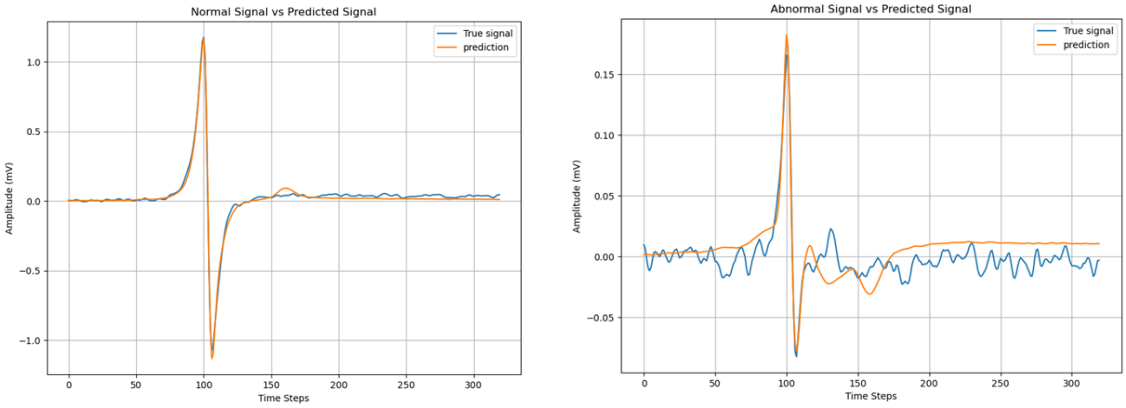
(a) Autoencoder Architecture One

(b) Autoencoder Architecture Two

FIGURE 12. Autoencoder Architectures: First architecture on the left has only one hidden layer with 5 latent representations; Second architecture on the right has three hidden layers with 20, 5, and 20 hidden units.

TABLE 7. Results from autoencoder for anomaly detection.

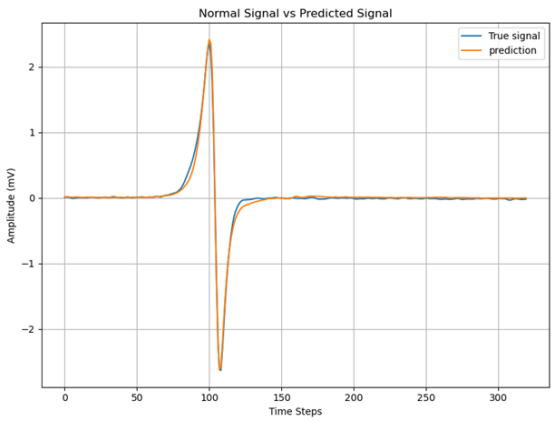
	Autoencoder with First Architecture	Autoencoder with Second Architecture
Relevant MSE for Training Set	0.106573	0.087316
Relevant MSE for Test Set	0.234638	0.232959



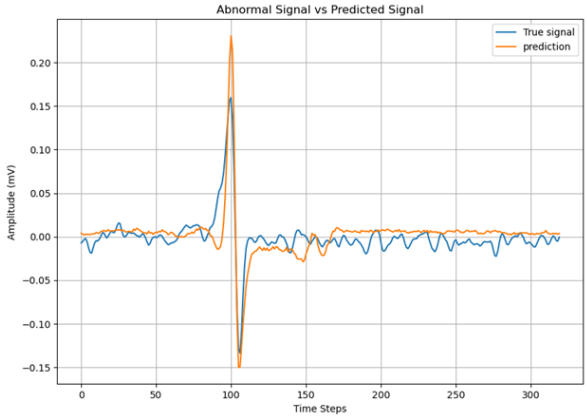
(a) An example of normal signal versus autoencoder reconstructed signal

(b) An example of normal signal versus autoencoder reconstructed signal

FIGURE 13. True Compared with Reconstructed Signals Using First Autoencoder: On the left, we have shown one example of the normal signal compared with autoencoder reconstructed signal; On the right, we have shown one example of the abnormal signal compared with autoencoder reconstructed signal.



(a) An example of normal signal versus autoencoder reconstructed signal



(b) An example of normal signal versus autoencoder reconstructed signal

FIGURE 14. True Compared with Reconstructed Signals Using Second Autoencoder: On the left, we have shown one example of the normal signal compared with autoencoder reconstructed signal; On the right, we have shown one example of the abnormal signal compared with autoencoder reconstructed signal.

BIBLIOGRAPHY

[1] Guns, P. J. (2020). "INSPIRE": A European training network in safety pharmacology creating opportunities for 15 PhD students. *Journal of Pharmacological and Toxicological Methods, 105, 106838.* https://doi.org/10.1016/j.vascn.2020.106838.

[2] Redfern, W. S., Wakefield, I. D., Prior, H., Pollard, C. E., Hammond, T. G., & Valentin, J. P. (2002). Safety pharmacology - a progressive approach. *Fundamental and Clinical Pharmacology.* 16(3), 161–173. https://doi.org/10.1046/j.1472-8206.2002.00098.x.

[3] Brake, Katherine & Gumireddy, Ashwini & Tiwari, Amit & Chauhan, Harsh & Kumari, Dunesh. (2017). In vivo Studies for Drug Development via Oral Delivery: Challenges, Animal Models and Techniques. Pharmaceutica Analytica Acta. 08. http://doi.org/10.4172/2153-2435.1000560.

[4] Chris Delaney, J. A., & Suissa, S. (2009). The case-crossover study design in pharmacoepidemiology. *Statistical Methods in Medical Research.* 18(1), 53–65. https://doi.org/10.1177/0962280208092346

[5] Meurs, K., Miller, M., & Slater, M. (1996). Comparison of the indirect oscillometric and direct arterial methods for blood pressure measurements in anesthetized dogs. *Journal of the American Animal Hospital Association.* 32(6), 471–475. https://doi.org/10.5326/15473317-32-6-471

[6] Recent advances in studies of cardiac structure and metabolism:vol. (1976). The metabolism of contraction. *American Heart Journal.* 92(3), 411. https://doi.org/10.1016/s0002-8703(76)80135-4

[7] Templeton, G. H., Platt, M. R., Willerson, J. T., & Weisfeldt, M. L. (1979). Influence of aging on left ventricular hemodynamics and stiffness in beagles. *Circulation Research.* 44(2), 189–194. https://doi.org/10.1161/01.res.44.2.189

[8] Hosseini, H. G., Luo, D., & Reynolds, K. J. (2006). The comparison of different feed forward neural network architectures for ECG signal diagnosis. *Medical engineering & physics*, 28(4), 372-378.

[9] Adams, E. R., & Choi, A. (2012). Using neural networks to predict cardiac arrhythmias. *2012 IEEE international conference on systems, man, and cybernetics (smc), IEEE.* pp. 402-407.

[10] Silipo, R., & Marchesi, C. (1998). Artificial neural networks for automatic ECG analysis. *IEEE transactions on signal processing*, 46(5), 1417-1425.

[11] Basile, A. O., Yahi, A., & Tatonetti, N. P. (2019). Artificial intelligence for drug toxicity and safety. *Trends in pharmacological sciences*, 40(9), 624-635.

[12] Shameer, K., Johnson, K. W., Glicksberg, B. S., Dudley, J. T., & Sengupta, P. P. (2018). Machine learning in cardiovascular medicine: are we there yet? *Heart*, 104(14), 1156-1164.

[13] Liang, F., & Liu, H. (2005). A closed-loop lumped parameter computational model for human cardiovascular system. *JSME International Journal Series C Mechanical Systems, Machine Elements and Manufacturing*, 48(4):484-493.

[14] Mirramezani, M., & Shadden, S. C. (2020). A distributed lumped parameter model of blood flow. *Annals of Biomedical Engineering*, 48(12):2870-2886.

[15] Moulton, M. J., Hong, B. D., & Secomb, T. W. (2017). Simulation of left ventricular dynamics using a low-order mathematical model. *Cardiovascular engineering and technology*, 8(4):480- 494.

[16] Segers, P., Stergiopulos, N., Verdonck, P., & Verhoeven, R. (1997). Assessment of distributed arterial network models. *Medical and Biological Engineering and Computing*, 35(6):729-736.

[17] Gul, R., Schuette, C., & Bernhard, S. (2016). *Mathematical modeling and sensitivity analysis of arterial anastomosis in arm arteries.* Appl. Math. Model. 40, 7724–7738.

[18] Gul, R., & Bernhard, S. (2017). *Optimal measurement locations for diagnosis of aortic stenosis and aneurysms in a lumped parameter model of systemic circulation using sensitivity analysis.* Int. J. Biomath. 10, 175116

[19] Yu, Y. C., Boston, J., Simaan, M., & Antaki, J. (1998). Estimation of systemic vascular bed features for artificial heart control, *IEEE Trans. Automat. Contr.*, vol. 43, pp. 765–777.

[20] Markert, M., Trautmann, T., Krause, F., Cioaga, M., Mouriot, S., Wetzel, M., & Guth, B. D. (2018). A new telemetry-based system for assessing cardiovascular function in group-housed large animals. Taking the 3Rs to a new level with the evaluation of remote measurement via cloud data transmission. *Journal of Pharmacological and Toxicological Methods*, 93, 90–97. `https://doi.org/10.1016/j.vascn.2018.03.006`.

[21] Flaten, M. A., Simonsen, T., & Olsen, H. (1999). Drug-Related Information Generates Placebo and Nocebo Responses That Modify the Drug Response. *Psychosomatic Medicine, 61(2), 250–255.* `https://doi.org/10.1097/00006842-199903000-00018`.

[22] Fetics, B., Nevo, E., Chen-Huan Chen, & Kass, D. (1999). Parametric model derivation of transfer function for noninvasive estimation of aortic pressure by radial tonometry. *IEEE Transactions on Biomedical Engineering.* 46(6), 698–706. `https://doi.org/10.1109/10.764946`

[23] Le, V. P., Kovacs, A., & Wagenseil, J. E. (2012). Measuring Left Ventricular Pressure in Late Embryonic and Neonatal Mice. *Journal of Visualized Experiments.* 60. `https://doi.org/10.3791/3756`

[24] Singh, K. P., Bartolucci, A., & Bae, S. (2015). *Introduction to Statistical Analysis of Laboratory Data* (1st ed.). Wiley.

[25] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering. *ACM Computing Surveys.* 31(3). 264–323. `https://doi.org/10.1145/331499.331504`.

[26] MacKay, David (2003). Information Theory, Inference and Learning Algorithms. *Cambridge University Press.* pp. 284–292. ISBN 978-0-521-64298-9. MR 2012999.

[27] Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology*), 63(2), 411-423.

[28] Davidson, I., & Satyanarayana, A. Speeding up k-means clustering by bootstrap averaging. IEEE data mining workshop on clustering large data sets (2003).

[29] Hawkins, S., He, H., Williams, G.J., & Baxter, R.A. (2002). *Outlier detection using replicator neural networks.* Kambayashi, Y., Winiwarter, W., Arikawa, M. (eds.) DaWaK 2002. *Springer, Heidelberg.* Lecture Notes in Computer Science, vol. 2454, pp. 170–180.

[30] Karpinski, M., Khoma, V., Dudvkevych, V., Khoma, Y., & Sabodashko, D. (2018). *Autoencoder neural*

networks for outlier correction in ECG-based biometric identification. 2018 IEEE 4th international symposium on wireless systems within the international conferences on intelligent data acquisition and advanced computing systems (IDAACS-SWS), pp. 210-215.

[31] Bisong, E. (2019). *Building machine learning and deep learning models on Google cloud platform*, pp. 59-64, Berkeley, CA, USA: Apress.

[32] Yan, H., Jiang, Y., Zheng, J., Peng, C., & Li, Q. (2006). *A multilayer perceptron-based medical decision support system for heart disease diagnosis, Expert Systems with Applications. Elsevier,* vol. 30, no. 2, pp. 272-281.

[33] Raju, V. G., Lakshmi, K. P., Jain, V. M., Kalidindi, A., & Padma, V. (2020). *Study the influence of normalization/transformation process on the accuracy of supervised classification. IEEE, 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*. pp. 729-735.

[34] de Carvalho, A.C.P.L.F.& Freitas, A.A. (2009). *A Tutorial on Multi-label Classification Techniques. In: Abraham, A., Hassanien, AE., Snášel, V.* (Eds). Foundations of Computational Intelligence vol 5. Studies in Computational Intelligence, vol 205. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-01536-6_8

[35] Thiriet, M. & Parker, K. H. (2009). Physiology and pathology of the cardiovascular system: a physical perspective. In L. Formaggia, A. Quarteroni & A. Veneziani (Eds.). *Cardiovascular Mathematics: Modeling and simulation of the circulatory system* (pp. 1-45). Milano: Springer. InTech. doi: https://doi.org/10.1007/978-88-470-1152-6

[36] Lafta, M. & Hassanain, A. (2008). Engineering Modeling of Human Cardiovascular System. In AlNahrain University, *College of Engineering Journal - NUCEJ*, 11, 307-314.

[37] Suga, H. & Sagawa, K. (1974).Instantaneous pressure-volume relationships and their ratio in the excised, supported canine left ventricle, *Circ Res*, vol. 35, no. 1, pp. 117–126.

[38] Dorf, R. & A. Svoboda, J. (2006). *Introduction to Electric Circuits* (7th Edition), John Wiley.

[39] (2017, October 25). *Integration and ODEs: scipy.integrate.odeint.* SciPy v1.0.0 Reference Guide. https://pageperso.lis-lab.fr/~francois.denis/IAAM1/scipy-html-1.0.0/generated/scipy.integrate.odeint.html

[40] (2005, September 23). Hindmarsh, A. C., and Petzold, L. R. *LSODA, Ordinary Differential Equation Solver for Stiff or Non-Stiff System.* NEA. http://www.nea.fr/abs/html/uscd1227.html.

[41] Hansen, N. (2016). *The CMA Evolution Strategy: A Tutorial.* arXiv. https://arxiv.org/abs/1604.00772

[42] Jun, T. J., Nguyen, H. M., Kang, D., Kim, D., Kim, D., & Kim, Y. H. (2018). *ECG arrhythmia classification using a 2-D convolutional neural network.* arXiv preprint arXiv:1804.06812.

[43] Pereira, J., & Silveira, M. (2019). *Unsupervised representation learning and anomaly detection in ECG sequences. International Journal of Data Mining and Bioinformatics,* 22(4), 389-407.

[44] Dau, H. A., Ciesielski, V., & Song, A. (2014). *Anomaly detection using replicator neural networks trained on examples of one class. Springer, Cham. Asia-Pacific Conference on Simulated Evolution and Learning*, pp. 311-322.

[45] 1. Lacroix P., & Provost D. (2000). Basic safety pharmacology: the cardiovascular system. *Therapie.* 55(1):63-9.

[46] Hezzell, M.J., Humm, K., Dennis, S.G., Agee, L. & Boswood, A. (2013), Relationships between heart rate and age, bodyweight and breed in 10849 dogs. *J Small Anim Pract*, 54: 318-324. https://doi.org/10.1111/jsap.12079.

[47] Strasser A., Simunek M., Seiser M., Hofecker G. (1997). Age-dependent changes in cardiovascular and metabolic responses to exercise in beagle dogs. *Zentralbl Veterinarmed A.* 44(8):449-60. https://doi.org/10.1111/j.1439-0442.1997.tb01130.x.

[48] Cuomo F., Roccabianca S., Dillon-Murphy D., Xiao N., Humphrey J.D, & Figueroa C.A. (2017). Effects of age-associated regional changes in aortic stiffness on human hemodynamics revealed by computational modeling. *PLoS* One. 2;12(3):e0173177. https://doi.org/10.1371/journal.pone.0173177.

[49] (2014, August 14). Devona Branch. EKG Interpretation. *SlideServe, 16*, https://www.slideserve.com/devona/ekg-interpretation.