Extraction of dynamical patterns from fluorescence microscopy images using recurrent neural networks

A DISSERTATION PRESENTED
BY
ANDRÉS OSPINA MESA
TO
THE DEPARTMENT OF SCHOOL OF ENGINEERING (URV)
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER'S DEGREE IN COMPUTATIONAL ENGINEERING AND MATHEMATICS
IN THE SUBJECT OF
ARTIFICIAL INTELLIGENCE(AI)

UNIVERSITAT ROVIRA I VIRGILI
TARRAGONA, CATALUNYA
&
UNIVERSITAT OBERTA DE CATALUNYA
BARCELONA, CATALUNYA
JUN 2021

Thesis advisor: Professor Raul Benítez Iglesias Andrés Ospina Mesa

Extraction of dynamical patterns from fluorescence microscopy images using recurrent neural networks

Abstract

Heart beating it's what keeps human living, the correct function of it, among other things, determinate the life quality, longevity and diseases appearing, what is known is that arrhythmia it's the most common disease in the cardiac system, what it's not so common is that can be detected earlier, it can exist and even not be visible in electrocardiogram, that's because the heart contractions occur at cellular level, then at tissue level and finally at muscular level, if the arrhythmia test it's done at muscular level, that means that can be observed also at cellular and tissue level, the objective of this work it is using Recurrent Neural Networks(RNN), more precisely, Autoencoders + LSTM, to identify anomalies, which can be arrhythmia or other diseases, based on electrophysiological signals at cellular level, extracted from patients' cardiac tissue, creating interactive data visualization such as dashboard in Hypertext Markup Language(HTML) format.

Contents

Listing of figures

Acknowledgments

Firs of all I would like to thank God and Holy Spirit, for illuminate my decisions, for made possible that I choose this terrific master, the people who helped me to organize the ideas, making the writing clearer and concrete, I would like to make a special mention to my parents, who encourage me to study every day, providing me huge aids, with the intention of making things simpler to me and looking for me to concentrate all my efforts in this thesis, also professor Raul, who was always available for answering questions, providing some space in his agenda in times that suit in my lunch schedule, God bless all the people who in one or other way made this thesis possible, I'll be eternally grateful.

1

Introduction

1.1 Context of Justification

Heart is the engine of almost all living beings, it's responsible for bombing blood to body extremities, carrying all nutrients and oxygen, though blood vessels, which belong to circulatory system, the heart in mammals consist of artria and ventricles, first one are located in the upper part of the heart, called atria, which at the same time subdivide into right atrium and left atrium, which receive the blood from variety of veins and arteries, the lower part, called ventricles, subdivided into two parts too, the right one, is in charge to pump blood to the pulmonary system for pulmonary circulation, where the blood is oxygenized and then received by the left atrium and send it to left ventricle, for being pumped to the whole body, then the blood return to right atrium, and cycle start over, this cardiac process generates two main muscle actions, when blood is received by the atrium, the muscle is relaxed and enlarged to allow blood entrance, this is called **diastole**, when blood need to be sent for oxygenation at the pulmonary system or needs to be sent to the whole body, the heart perform a muscle contraction called **systole**, that make heart volume to decrease and forces blood to move for the open veins or arteries, when this contraction and relaxation does not work correctly, it's called Arrhythmia. Arrhythmia is one of the most common heart diseases in humans, particularly atrial fibrillation (AF) is the most common kind of Arrhythmia, that affects between 1-2% of population, the genesis of AF requires an atrial substrate, which formation depends on variety of factors, including lifestyle, genetic predisposition, blood pressure, obesity, diabetes, smoking and others, there are multiple factors that can be modified and other that not[9]. Deepen into Heart contraction, it's found that heart it's composed by multiple muscular tissues, that allow it to contract and relax, these tissues are formed for multiple cells, named **myocytes**, which the main characteristic is its ability to contract, specifically reduce their size[6], when the cardiac tissues contract by effect of cardiac myocytes contraction, the heart is contracted also, that's called a beat. The cellular contraction is produced by a process called "Excitation-Contraction Coupling", the peacemaker cells, these are myocytes modified cells, that does not have the ability to contract, however are specialized in generating and conducting the action potentials (electric pulses), these potentials are generated spontaneously, the main concentration of this peacemaker cells is the *SA node*, which is located in the upper part of the right atrium, the cardiac myocytes are full connected through gap junction, that allows action potential trigger another action potential of the adjacent cell, making the pulse propagate rapidly; All this action electric pulses belong to a polarization and depolarization of the pacemaker cells, allowing a transfer of ions, mainly calcium (Ca^{2+}), potassium (K^{+}) and sodium (Na^{+}), between the intracellular fluid (ICF) and extracellular fluid (ECF), the ICF is main of the time negative polarized and ECF more negative polarized, but for the purpose of explanation, let's say it's positive polarized, besides that, the outside of the cell (ECF) has more presence of sodium (Na^{+}) and calcium (Ca^{2+}), in the other hand, inside the cell (ICF) are more presence of potassium (K^{+}), the communication between both sides is by channels, one for each kind of ion, those who open and close in certain moments to keep the state of the cells, the spontaneous action potential, conclude in the opening of some of these channels allowing ions come into the cell and other ions to get expelled outside, in all of this process the polarization decrease and increase, when the potential triggers the cardiac myocytes channels, that in fact has different intracellular and extracellular ion's composition, which is formed by more potassium inside the cell and more calcium and sodium outside, moreover the cells has a resting potential of $-90mV$, when peacemaker or other previous activated cell potential arrives, it produces a depolarization, due to this voltage change, the sodium channel is opened, and intracellular fluid is full of this sodium ions, then the calcium channel is opened too and ions go from ECF to ICF, the potassium channels opens and ions go outside the cell, the calcium is actively transported it to outside the cell and that the contraction phase[8]. As was introduced the contraction produced by the electrical stimulation is fundamental for heart beating, so for living in consequence, the main motivation behind is improving people lifestyle and early heart disease detection, when an arrhythmia is detected by medics with help of electrocardiograms, sometimes the disease have been present for year, but was not possible to detect with this method, because anomalies were imperceptible for human eye and even for electrocardiograph, when an anomaly is detected was due to an anomaly beating at cellular level, if the analysis begins at that level, diseases could be detected earlier and treatment also could be more effectively, with laboratory experiments, is possible to replicate the heart beating at tissue level and also measure the pulses, and then use Artificial Intelligence(AI) to detect anomaly pulses, training a Recurrent Neural Network (RNN) with a normal (no anomaly) pulses and then pass it an anomaly pulses signal, with hope that the RNN detects and mark the anomalies, making easier for medics and researchers identify the anomalies and treat the patient in the early phase, increasing the response of the body to treatment and making process simpler and less intrusive for patient.

1.2 Objectives

The main objective for the work, it's to collect, organize and prepare cardiac myocyte pulse signals from experimental simulation with external stimulation, both anomaly pulses and normal ones.

- Create an architecture capable to learn the data features and behavior so well that detects the anomaly presence in any part of cardiac cells signals.
- Train a Recurrent Neural Network(RNN) sufficient number of signals and test it, it's expected that the system could accomplish an acceptable rate of detections.
- For making this research accomplishments accessible to concrete solution, it's planned to create a tool that can be used in the field, with real patients and letting medics use it for their daily work, carrying the academic research to the practical daily tasks, for this purpose, the tool will be developed such the user can pass a normal signal or also none, and the patient with potential heart disease signal, such as return the patient signal with anomalies identified with red color in case there exists and other informative charts for giving the user a better understanding of the patient beat behavior, the end user result will be displayed in HTML file, that can be opened in many browsers, making more accessible for general users.

1.3 Approaches and Method

Making a brainstorm about what would be the best strategy to tackle the problem, 2 main ideas were preselected, first one was to train a neural network with so many samples of normal behavior signals, with the intention of it be capable to recognize any anomaly when the suspicious signal were provided, second one was to create an architecture that could recognize the anomalies within a signal based on another normal behavior signal of the same kind, this approach will always require 2 signals, one for the training of the network and other one for testing (the evaluation signal), both approaches considered an application where the user can do its test, use the model with autonomy and simplicity, considering that the end user it's not expert in computer science field;Finally the second approach was selected, based on the better results that could have, when the training is made with a specific reference of the signal that are interested in evaluate, the learning is more precise, and take to fewer confusions with other signals, moreover the first method is more practice and require only one signal, but for the developing phase demand many more signals for training, include at least ten samples of each kind of signal from each channel, and also an incredible computer power, that would require incurring in high costs. To conclude, the approach for this work is about the developing a new product, based on previous investigation about the heart diseases made by researchers in fields like biomedical, computer science and vet.

1.4 Work Planning

For making the explanation clear and chronologically organized, the task and milestones will be itemized below.

1.4.1 Task

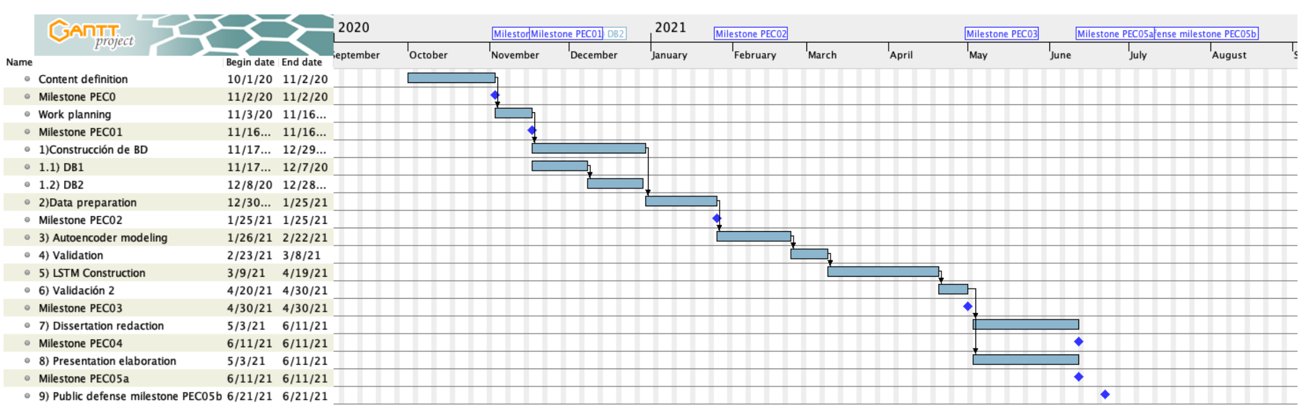
I Database construction, starting from initial data coming from electro-physiological signals of cardiac myocytes.

  I Database construction, which would have calcium (Ca^{2+}) response signals, with both normal (healthy) and anomalous (ill) behavior, containing different types of anomalies, these signals could be obtained, modifying normal signals or anomalous one, this signals modification will be called "realistic synthetic signals", based on these signals it's expected that the network can learn some different kinds of anomalies and detect them afterwards.

II Data preparation and preprocessing, starting from the database constructed previously, it is processed in such way that information which will be analyzed afterwards, add the most explicability and fewer confusions, all this will consist in: noise filtering, eliminating trend and any kind of procedure needed to accomplish this task objective.

III Modeling part, consist on the construction of recurrent neural network autoencoder architecture, find the best parameters that optimize its results, in such way that the loss function be optimal, and show the goodness-of-fit of the model.

IV Validation of results, in a visual way, it is know where are the anomalies, and it's possible to determinate if the model detected well or not each one.



1.4.2 Milestones

I PEC0 Content definition.

II PEC01 Work planning.

III PEC02 Developing part 1.

IV PEC03 Developing part 2.

V PEC04 Dissertation redaction.

VI PEC05a Presentation elaboration.

VII PEC05b Public defense.

1.5 Obtained products

After work is completed, the products that will be obtained must be: a recurrent neural network architecture, particularly Autoencoder + LSTM, with its parameters optimized and performance validation, the evidence of the validation and the explanation of what is seen in each one, lastly an application where the end user can introduce signals for testing and get and HTML file with interactive charts, allowing the use of certain tools for deepen analysis or explanation to peers in other contexts.

1.6 Chapters description

When introduction of the work is done, and the concepts that will be developing during dissertation are clear, it's important to remark the order in which will be developed each milestone mentioned previously, this description intention is to clarify what would be found in each chapter and what are the concepts that would be expanded, moreover the chapter follows the Gantt diagram presented above. First chapter is the data extraction, which explain how the data was collected, how the experiment was constructed and what data makes reference to, the importance of getting good data quality and the organization of signals, that would be more relevant in the modeling chapter, for coding and understand what signal is being analyzed, just by its name, this chapter also talks about the nature of myocytes cells, how the contraction works and the repercussion of this in correct heart function. Second, the recurrent neural network architecture would be explained, what artificial intelligence (AI) is, why this method is a good approach to solve this kind of problems, in the same way the architecture and data flow is detailed, making easier to abstract the information and transform it into a logic flow that can be easier understood, the LSTM(Long Short Term Memory) contribution to give context to the problem and its great combination with autoencoder, showing better results than autoencoder by its own, this section is more illustrative about the method that will be used, in addition clarify the challenge of treating this kind of data with unsupervised learning. Third, the modeling chapter is the most extensive, particularly because describe step by steps what was done with data, before, during and after the data passing through network architecture, further the explanation of why each step it's needed and how that process helps to obtain more consistent results, this chapter is very illustrative and is the coding part of the entire work. Fourth, Results presentation, this section is the most satisfying, first because it's the final product after all previous chapters, and also it's a tangible product, that can be used for many other purposes and shows how all the components that sometimes seemed ethereal, becomes something concrete, this section presents one by one, each of the signals that were evaluated, a figure that shows the anomalies detected, the original signal image that was tested, the signal that was used to train the network, the *MAE* chart, where it overpasses the threshold line and consecutively can be marked as anomaly if last longer than 95 data points of 190 in the current window. Ultimately, the conclusions chapter, talks about what was learned with all this developing, what can be extracted for any chapter of the work, and it is important to remember or to bring up, it is a reflection about what was done, the advantage of the final results and possibles continuations that can the work have, it also shows the challenge that appeared during the elaboration and how were solved, this is essential for other authors that would like to base their projects or take a reference.

"Without ambition, one starts nothing. Without work, one finishes nothing. The prize will not be sent to you. You have to win it."

Ralph Waldo Emerson

2

Experimental Data

The data of signal pulses were obtained thank to Dr. Marcel Jiménez Farrerons, who has been working with heart arrhythmias research for long time, and has directed thesis about this subject, particularly the data used for this work was extracted from laboratory experimental setup, during develop of

Veterinary degree thesis[3],in the experiment was used a OF1 mice, who was anesthetized and confirmed that was correctly done, Second was euthanized by atlantooccipital dislocation, to obtain the atrium, two incisions were made in the torax cavity, making possible the heart extraction and lungs separations, with heart extracted and isolated, a microscopy was used to identify were was the right and the left atrium, then were tied to silk and placed in the setup constructed for the experiment, this setup 2.1 consisted in the atrium tied to a force transducer, which will measure the movement of each tissue, each were placed separativly, introduced in salinized solution with constant temperature of 37.5° C, then with help of impulse generator, the left artia was stimulated, because as was exposed previously, the right artia has its own stimulation (peacemaker cells) and the response were stored by powerlab/800 software.

The experiments had multiple variations, the date when were made, the channel used in each experiment and the artia used, this data was stored with following name format:

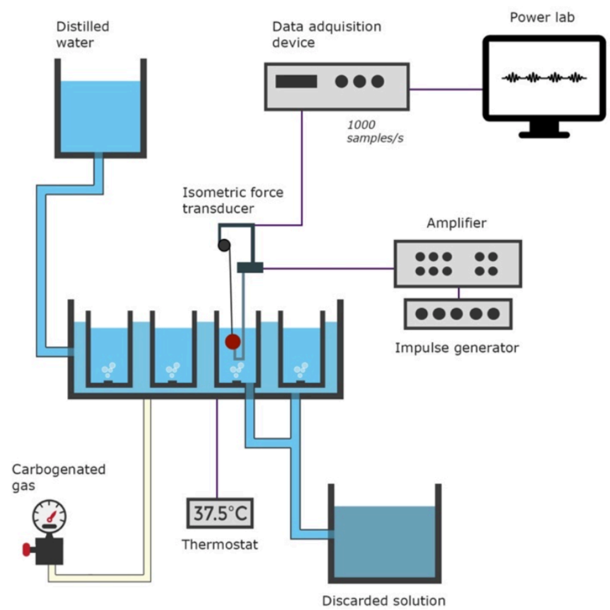


Figure 2.1: Experiment set-up, adopted from [3] p. 5

"MM-dd-aa-a(i-d)-c(1,2,3,4,5,6)-6hz-(a or n).txt", which means the month, the day, and the year when the experiment was performed, then which artia was used, "i" meaning left artia and "d" meaning right artia, consecutively the channel used, which goes from 1 to 6, then for all experiments was used 6hz frequency, lastly the label for indicate if it was a normal(n) response, or it had an anomaly(a) pattern, all this data were stored in text(.txt) format. This data were reserved for more advanced stage of the work. For starting the data exploration and having some samples to start the developing phase, it was used one

2.2 of 3 signals that illustrate a normal behavior of a pulse, from this data, it was constructed 6 realistic syntetic signlas that map most common anomalies that has presence in patients, according to thesis advisor: Dr. Raul Benítez Iglesias and Dr. Marcel Jiménez Farrerons.

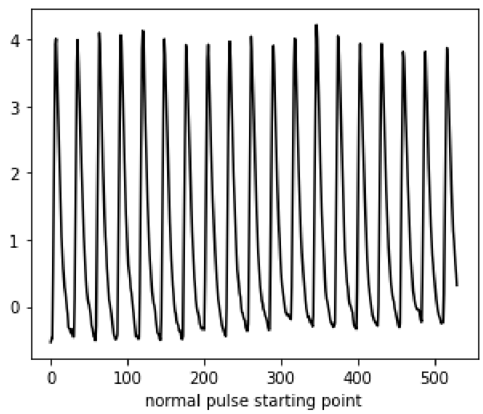


Figure 2.2: Baseline normal signal

for obtain this realistic synthetic(Baseline images modified) images 2.3, it's needed data processing, and modify the data values and also the time-lapse of some beats, for this objective, it was used the programming language(PL) Python in its version 3.7.3 with the Interface Developing Environment(IDE), Google Colaboratory, that is a structure of notebook with the advantage of GPU provided by Google, that will be awfully later in training and testing phase, using the libraries pandas[7], numpy[4] and scipy[10], the signal data is read, transformed into dataframe and finaly modified, giving as result 6 signals, that at all represent the anomalies that are of interest in the study.

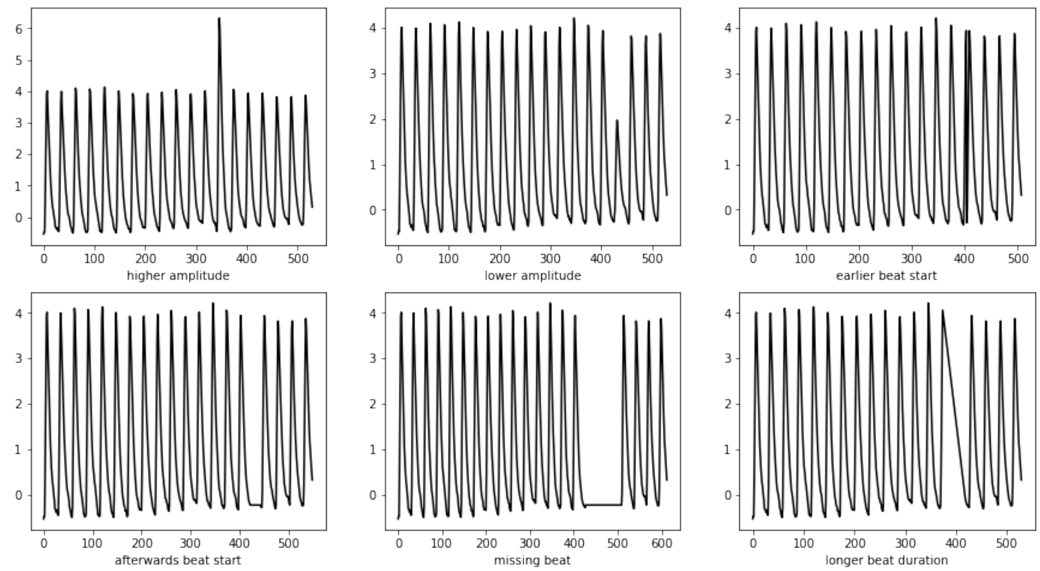


Figure 2.3: Realistic synthetic anomaly signals

"If you tell the truth you don't have to remember anything."

Mark Twain

3

Anomaly Detection Using Recurrent Neural Networks

Artificial intelligence is the capability that machines can develop in order to emulate the natural intelligence of humans, is not to forget that humans are not the only whose has intelligence, most of the mammals have any kind of intelligence, but humans are those who has the more developed intelligence so far, moreover the machines can develop certain kind of intelligence, like take decision based on data, recognize objects, learn and also teach, all of this has been developed with help of algorithms and data, but it's not as new as people think, this area started around 1950s, but this time has been more relevant because of the growing of computational capability, which is one of the main tools required to develop any kind of training, machines needs a lot of data and many examples to learn to do something by its own, this is called supervised learning, when the computer has data that defines an output, in the other hand, when the computer doesn't know the output, it needs a specialized algorithm to interpret the data and get conclusion that can be applied in future data, this kind of learning as is expected it's less accurate and demand better data quality and many more examples, this last type of learning will be used for the anomaly detection purpose in cardiac myocythes pulses, it's the most accurate kind of learning, based on the idea that the signals does not have any labels that identifies their anomalies precisely, the data is a collection of points and are analyzed by medics and researchers, that based on the experience and biological studies, declare that each particular signals represents a normal behavior or has any kind of anomaly that could or not be registered in the science knowledge.

Autoencoders is a neural network, that belongs to the unsupervised learning category, and its main function it's to learn from input data, extract the main features, also knows as characteristics, then reduce the data dimensionality, finally reproduce input data based on the reduced dimensionality data, which in process focus on reduce the loss function, that in most cases is the mean absolute error (MAE), $MAE(y, \hat{y}) = \frac{1}{m}\sum_{i=1}^{n}|y_i - \hat{y}_i|$ which in this case is the difference between the output reconstructed and the input data, with which was trained the network, as can be deduced, the autoencoder's architecture consist in general of three parts: Encoder, Code and Decoder 3.1

So far this is the starting point for modeling the problem and propose a solution, even though autoencoder by its own, it's not as accurate as needed

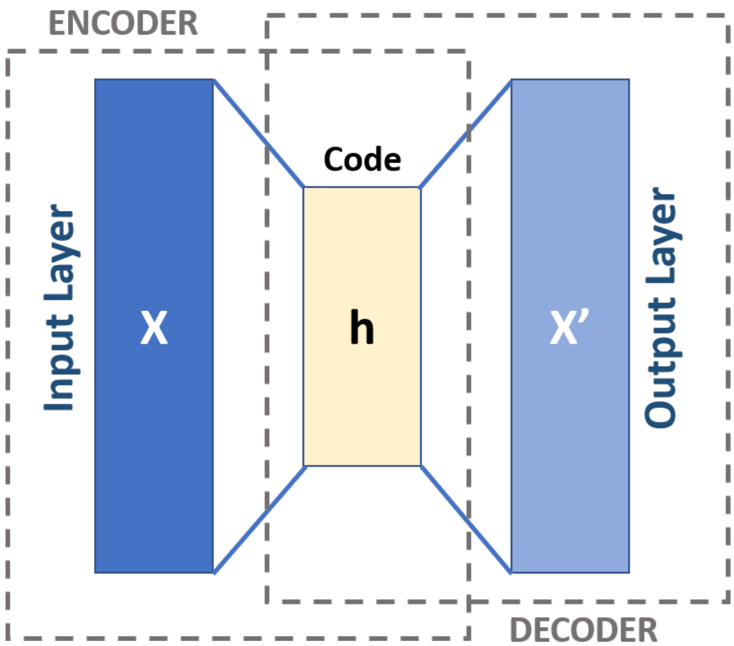


Figure 3.1: Autoencoder architecture
[2]

for signal processing, this kind of problems are denominated as sequence-to-sequence problems, where the time and the order of the data is important for the next data point, and sometimes the events happens at specific point, making more complex the learning for traditional RNN, for all previous reasons, Long-Short-Term Memory(LSTM) architecture is included into solution, the LSTM goodness it's that include the "context" of the problem as feature for learning and predicting next data points, that means that it's more similar as humans think, use the context of a problem for solving next task, the old events take less importance, that recent events, that why this architecture has a series of gates, that in most traditional case are three: "forget gate layer", "input gate layer" and the "output gate" 3.2, first one is responsible for deciding what information keep and what information discard, all this based on a sigmoid function, which has values between 0,1, where 0 represents a discard decision and 1 represents a remember one, secondly the input layer, is responsible for include data that will be remembered, it updates some values present in the network, and are updated with new data or is completely new data that is inserted into the network, all this based on a sigmoid and tanh function, finally the output that decide what to present in the next data point, based in the new state of the memory,

This addition to autoencoder, helps to identify the important patterns of the signals, in our case and each time the beat occurs might be updated that value, gaining relevance again, then when the prediction time arrive, that beat

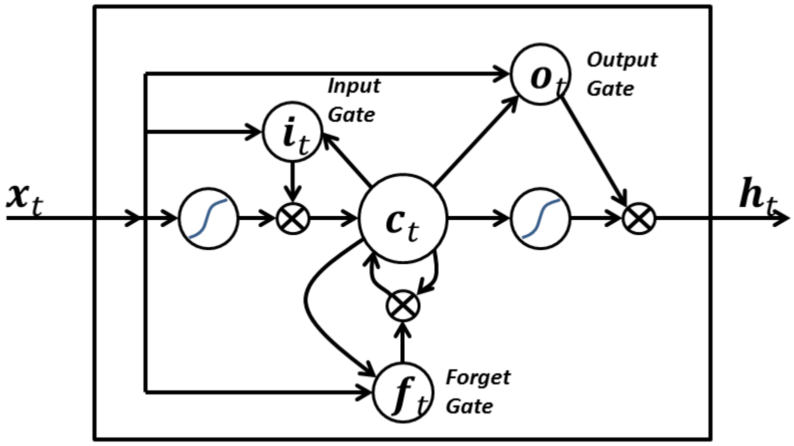


Figure 3.2: LSTM architecture

will be predicted and in case the test signal has that beat missing, the error will be greater, and then it will be detected as an anomaly; for building this Architecture of Autoencoder + LSTM, it was used the library Keras with python language, this architecture is the following

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_12 (LSTM)               (None, 190, 100)          40800
_________________________________________________________________
lstm_13 (LSTM)               (None, 64)                42240
_________________________________________________________________
dropout_6 (Dropout)          (None, 64)                0
_________________________________________________________________
repeat_vector_3 (RepeatVecto (None, 190, 64)           0
_________________________________________________________________
dropout_7 (Dropout)          (None, 190, 64)           0
_________________________________________________________________
lstm_14 (LSTM)               (None, 190, 64)           33024
_________________________________________________________________
lstm_15 (LSTM)               (None, 190, 100)          66000
_________________________________________________________________
time_distributed_3 (TimeDist (None, 190, 1)            101
=================================================================
```

Figure 3.3: Model architecture

"A strong, successful man is not the victim of his environment. He creates favorable conditions. His own inherent force and energy compel things to turn out as he desires."

Orison Swett Marden

4

Generation of Synthetic Data and Validation of the Approach

For starting this part of the work, it is important to remark that the modeling will be developed in multiple steps, which are the process for most of the data based projects, First, data processing is required for getting a valid statistical estimator, and guarantee that the data is the mode that RNN requires for their training and testing, due to this initial, but crucial step, the data of each signal it's standardized, which means that the data probability distribution fulfill a normal distribution, it means a Gaussian distribution with mean equal zero and unit variance, this is extremely recommended when working with data and time series concretely, using the function StandardScaler inside the preprocessing mode of sklearn library, it is accomplished successfully, second, the data is reshaped into only one column data, making easier other operations later in the modeling, keeping in mind that processes already described are applied for both training and testing data; for training the LSTM recurrent neural network and also autoencoder, besides of passing the entire signal data, it's better practice and had shown better results in the state of the art, to create sequence of the signal, that means that the number of data points will be greater, which for Recurrent neural networks (RNN) training is always better and also make the network more robust, because will be trained with many more samples and will get more "experience", similar to human brain learning, for this sequence it's define a window size, in the code is named as *TIME_STEPS*, this will define the number of data points that will have each sequence, with this *TIME_STEPS* size set to 190, this number was set based on lots of testing, with many values and evaluating the anomaly detection performance with each one, after that parameter set, then sequences are created from *TIME_STEPS* to then length of the signal minus *TIME_STEPS*, extracting a first segment of the signal with size 190, with 190 as starting point and 380 as ending point, then the second segment of the sequence will be from 191 to 382, and so on, all these parts are stacked into an array, next the array of sequences is reshaped like a tensor, in order to be trained by the network, in this case will be a tensor of 3 axes, composed by length of signal for 1-axis, 190 (TIME_STEPS) for 2-axis and 1 for 3-axis, the architecture chosen is then modeled with keras functions and compiled with Adam optimizer, which it's a stochastic gradient descent method and based on adaptive estimation of first and second order[5], this compilation on also uses a learning rate of 0.001, this means how fast or slow does the network change its weights, for adapting to the problem, this value goes from 0.0 to 1.0, a smaller value will require more epoch to converge, in this case the MAE loss to converge, a high value for the learning rate will take to an early convergence, the network will take the *MAE* as loss function as minimization objective as was barely mentioned before; when the acutoencoder LSTM architecture is configured and compiled and also the data is preprocessed, it's time for starting the training phase, at this point it's needed to set some parameters that will determinate the training behavior, there are a lot of parameters than can be used, but in this case were used: 1. Number of epoch, 2. Batch size, 3. Validation split 4. One callback and 5. Verbose during process.

let's detail each one, first the number of epoch makes reference to the number of loop or iterations that the training process will make, it means that 100 epoch, which it's the case in the project, data is going to pass through the network 100 times, and the network will adjust its weights probably the same number of epochs, the more number of epoch the more accurate will be the network, but it doesn't mean that every epoch will reduce the loss function, that's going to be explained in more detail later, second, the batch size is related with epoch in the way that is used for training the network by steps, this batch size refers to the number of samples that will be propagated through the network, in other words, the RNN needs to be trained with N samples from the training dataset, the batch size defines the number of this data points that will be passed at the same time, so the number of times that network will be trained is $\frac{N}{batch\_size}$, when all dataset is completely covered, it counts as 1 epoch, as the epoch, the batch size has interference in the training time and convergence quality, as larger is the batch size as faster will be the network training and vice versa, third, validation split it is a portion of the training dataset, that will be reserved for validating the loss function and by consequence adjust its weights to get predictions as similar as validation data samples, forth, a callback it's by definition an object that could perform actions at certain stages before, after or during the training phase, in this case, it's used a callback called *EarlyStoping*, it performs a stop training action, when defined conditions are satisfied, concretely when the loss function optimization does not improve after the number of epoch set in another parameter named *patience*, which is 15 for this work, finally, verbose is an optional parameter that has not interference in the training process or in its results, it's merely for get a cleaner output when code in execute, when this parameter it's set to 0, the training process will not generate any output in the console, this configuration is completely personal and depends on the developer likes. Given the above, the predicting phase has arrived, the already trained network is used for predicting, first the training data is predicted and is used for getting an important metric, train_mae_loss, which it's the *MAE* between the training data and the predicted training data, this result is used for establishing a limit or better called threshold, the idea it's to have a value that can be used as criteria for discern between a normal data point or an anomaly one, for getting this threshold, it's applied the mean of 95th percentile plus 5th percentile, ultimately the test data from 6 signals (realistic synthetic data), is pass it through the network with the intent of localizing the anomaly beats, so same process is following as for training process, then a sequence of data is created based on the testing samples, second, the stacked array of sequences is reshaped and the test_mae_loss function is calculated, with this previous metric it's time for comparing each value of the test_mae_loss array with the threshold calculated previously, when a value from test_mae_loss array is less than the threshold, it's stored into another array called *anomalies*, with a "FALSE" value, representing that it's not an anomaly data point, else with "TRUE" value, representing that is an anomaly, but this is not enough for localizing the anomalies, it is important to remember the LSTM architecture purpose, giving context to the problem-solving, in this case was defined to create sequences of points to train the network not point by point but sequence by sequence, for accomplish this purpose, a python function was created that evaluates the *anomalies* array with a window function, this array is evaluated in a range, from 190-1 to length of test signal minus 190+1, this will be called "data_idx", so when the sum of "TRUE's" of

```
anomalies[data_idx - TIME_STEPS + 1 : data_idx]
```

is greater than $\frac{TIME\_STEPS}{2}$, then that particularly data point is confirmed as an anomaly, in fact the analysis for detecting this points that are out of threshold, are identified within a context of another 189 data pints, which allow including multiple decision-maker for each data point, this helps to exclude some false positives, that may be out threshold, but are isolated events or outliers, that no represents part of a complete anomaly beat, as result the training and testing phase is complete, a good illustration of these results, helped during work modeling for confirming that the architecture and the method followed was suitable for this arrhythmia problem at cellular level; The confirmation results will be presented below 4.1, it can be observed that the signals originally are plotted in black color, the anomaly part detected by the model will be plotted over the signal part with red color.

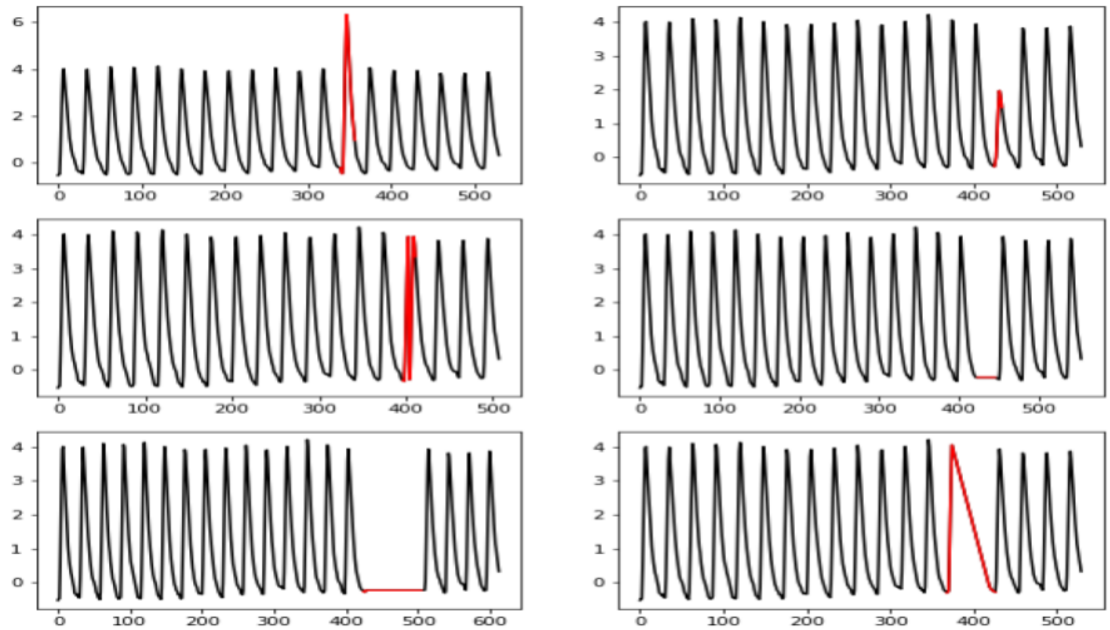


Figure 4.1: Anomaly detection over realistic syntetic signals.

To summarize, the training process was made with the signal 2.2, then each signal showed in the signal above 4.1, then the model predict based on this signal and plot the signal as was described above, concluding, 6 signals anomalies where predicted with a network trained with same signal, which is 2.2, the method chosen is validated and can be applied to real signals, despite this architecture and method test, the real signals could present different new challenges, that will need other approach or adjust some part of the method. Another part of the work, consist in an application where the end users, in this case researchers and medics, could use to evaluate their signals, and give them some tools to make deeper analysis on the signals, for accomplish this, a front-end library named *Plotly* is used, it allows creating Machine Learning (ML) applications, focused on data visualization with interactive charts, that allow actions such zoom in, zoom out, autoscale, pane, download as image, etc., also plotly allow deploying de app in multiple platforms, in basic or advanced integration, even though it was chosen the HTML export method, which is the most compatible format, can be used in any Operative system(OS), does not require many machine resources, can be used even without internet (offline mode) and has not direct cost of infrastructure, all this made it the best option for this particular project. Elaborating over this idea, for creating this HTML file, showing the results of the model, was used as for the model python language with plotly library, first it's needed to choose the figures that are important for analysis by medics in the heart disease researches, 4 figures were selected as relevant, 1. The original signal with normal pattern, with the preprocessing done for the modeling, to make the results comparable visually,

2. The signal that are interested to test, suspicious of having anomalies, also with preprocessing, 3. The RNN prediction, 4. The MAE loss function between the predicted signal and the test signal, combined with the threshold representation, this combination will show the positions where the *MAE* surpass the threshold line and in what measure, 5. Lastly the testing signal with its anomaly parts pointed out in red color, such images showed in the validation of the network. After these definitions are made, the developing part can start, a python function is created with the outputs of the model, which are more than the already mentioned, the outputs used as input for the application are: the training signal, the test signal, the predicted signal, MAE loss of training data, MAE loss of testing data, the anomaly indices, this last one used for locate the anomalies within the test signal, with all this data function is code, taking into account labeling all charts, using red color for anomalies and black for the rest, all charts has its titles and moreover the layout of the file is divided in two columns, left one for signal charts, and right one for histograms of training and testing MAE loss.

"Knowing is not enough; we must apply. Willing is not enough; we must do."

Johann Wolfgang von Goethe

5

Results (With Experimental Data)

Real signals provided by doctor Dr. Marcel Jiménez Farrerons, were divided in two groups, right atrium and left atrium, also each group contained such normal signals, such anomaly ones, the signals are listed in the following tables with its respective formatted name, first 2 tables will present the normal behavior signals (training signals), the next two tables will present their counterpart and other anomalous signals

Left atria		
Signal	Name	behavior
1	08-10-19-ai-c7-6hz-n.txt	Normal
2	11-10-19-ai-c1-6hz-n.txt	Normal
3	11-10-19-ai-c3-6hz-n.txt	Normal
4	13-02-20-ai-c4-6Hz-n.txt	Normal
5	14-02-20-ai-c3-6hz-n.txt	Normal
6	14-02-20-ai-c6-6Hz-n.txt	Normal
7	15-11-19-ai-c4-6hz-n.txt	Normal
8	15-11-19-ai-c5-6hz-n.txt	Normal
9	21-05-19-ai-c4-6hz-n.txt	Normal
10	21-05-19b-ai-c4-6hz-n.txt	Normal
11	28-05-19-ai-c1-6hz-n.txt	Normal
12	4-10-19-ai-c1-6hz-n.txt	Normal
13	4-10-19-ai-c2-6hz-n.txt	Normal
14	5-11-19-ai-c1-6hz-n.txt	Normal
15	5-11-19-ai-c3-6hz-n.txt	Normal
16	8-10-19-ai-c8-6hz-n.txt	Normal
17	9-10-19-ai-c7-6hz-n.txt	Normal
18	9-10-19-ai-c8-6hz-n.txt	Normal

Right atria		
Signal	Name	behavior
1	02-03-21-ad-c1-n.txt	Normal
2	05-11-19-ad-c5-n.txt	Normal
3	05-11-19-ad-c6-n.txt	Normal
4	07-10-19-ad-c3-n.txt	Normal
5	08-10-19-ad-c5-n.txt	Normal
6	10-10-19-ad-c5-n.txt	Normal
7	10-10-19-ad-c6-n.txt	Normal
8	10-10-19b-ad-c5-n.txt	Normal
9	11-10-19-ad-c5-n.txt	Normal
10	12-11-19-ad-c5-n.txt	Normal
11	15-11-19-ad-c1-n.txt	Normal
12	21-05-19-ad-c1-n.txt	Normal
13	24-02-21-ad-c3-n.txt	Normal
14	24-02-21-ad-c4-n.txt	Normal
15	25-02-21-ad-c3-n.txt	Normal
16	25-02-21-ad-c4-n.txt	Normal
17	26-03-21-ad-c2-n.txt	Normal
18	28-05-19-ad-c4-n.txt	Normal

Left atria		
Signal	Name	behavior
1	11-10-19-ai-c3-6hz-a.txt	Anomalous
2	13-02-20-ai-c1-6hz-a.txt	Anomalous
3	14-02-20-ai-c3-6hz-a.txt	Anomalous
4	14-02-20-ai-c6-6Hz-a.txt	Anomalous
5	15-11-19-ai-c4-6hz-a.txt	Anomalous
6	15-11-19-ai-c5-6hz-a.txt	Anomalous
7	15-11-19-ai-c5-6hz-a2.txt	Anomalous
8	15-11-19-ai-c5-6hz-a3.txt	Anomalous
9	15-11-19-ai-c5-6hz-a4.txt	Anomalous
10	15-11-19-ai-c5-6hz-a5.txt	Anomalous
11	4-10-19-ai-c1-6hz-a.txt	Anomalous
12	4-10-19-ai-c2-6hz-a.txt	Anomalous
13	5-11-19-ai-c1-6hz-a.txt	Anomalous
14	5-11-19-ai-c3-6hz-a.txt	Anomalous
15	6-3-19-ai-c2-10hz-a.txt	Anomalous
16	6-3-19-ai-c2-1hz-a.txt	Anomalous
17	6-3-19-ai-c2-2hz-a.txt	Anomalous
18	6-3-19-ai-c2-3hz-a.txt	Anomalous

Right atria		
Signal	Name	be
1	01-03-19-ad-c3-a.txt	Ano
2	02-03-21-ad-c2-a.txt	Ano
3	04-10-19-ad-c4-a.txt	Ano
4	08-10-19-ad-c5-a.txt	Ano
5	09-10-19-ad-c5-a.txt	Ano
6	09-10-19-ad-c6-a.txt	Ano
7	10-10-19-ad-c5-a.txt	Ano
8	10-10-19-ad-c6-a.txt	Ano
9	14-02-20-ad-c4-a.txt	Ano
10	14-02-20-ad-c6-a.txt	Ano
11	17-03-21-ad-c2-a.txt	Ano
12	17-06-19-10-ad-c4-a.txt	Ano
13	17-06-19-9-ad-c4-a.txt	Ano
14	21-05-19-ad-c1-a.txt	Ano
15	25-02-21-ad-c4-a.txt	Ano
16	25-03-21-ad-c1-a.txt	Ano
17	26-03-21-ad-c1-a.txt	Ano
18	28-05-19-ad-c4-a.txt	Ano

The number of samples for each atrium is enough for the analysis, but as was explained in the model part, it's needed to train the network with a normal signal of the same type that which will be evaluated, this means multiple factors, as the atrium selected, the channel of calcium, which was evaluated, etc., for his reasons it was chosen the same signals with counterpart for each atrium, given the above, the signals output are hosted in a GitHub page [1], each signal also points to the interactive visualization hosted

for the left atrium got these ones

Left atria signals evaluated	
Normal	Anomalous
4-10-19-ai-c1-6hz-n.txt	4-10-19-ai-c1-6hz-a.txt
4-10-19-ai-c2-6hz-n.txt	4-10-19-ai-c2-6hz-a.txt
5-11-19-ai-c1-6hz-n.txt	5-11-19-ai-c1-6hz-a.txt
5-11-19-ai-c3-6hz-n.txt	5-11-19-ai-c3-6hz-a.txt
9-10-19-ai-c8-6hz-n.txt	9-10-19-ai-c8-6hz-a.txt
11-10-19-ai-c3-6hz-n.txt	11-10-19-ai-c3-6hz-a.txt
14-02-20-ai-c3-6hz-n.txt	14-02-20-ai-c3-6hz-a.txt
14-02-20-ai-c6-6Hz-n.txt	14-02-20-ai-c6-6Hz-a.txt
15-11-19-ai-c4-6hz-n.txt	15-11-19-ai-c4-6hz-a.txt
15-11-19-ai-c5-6hz-n.txt	15-11-19-ai-c5-6hz-a.txt

same for the left atria

Right atria signals evaluated	
Normal	Anomalous
28-05-19-ad-c4-n.txt	28-05-19-ad-c4-a.txt
08-10-19-ad-c5-n.txt	08-10-19-ad-c5-a.txt
10-10-19-ad-c5-n.txt	10-10-19-ad-c5-a.txt
10-10-19-ad-c6-n.txt	10-10-19-ad-c6-a.txt
21-05-19-ad-c1-n.txt	21-05-19-ad-c1-a.txt
25-02-21-ad-c4-n.txt	25-02-21-ad-c4-a.txt

Before presenting the results, it's important to clarify that the left atrium signals were externally stimulated, whence the beats are more dynamics and in some cases has beats that seems clear for detection as anomaly, in the other hand the right atrium signals were not externally stimulated, due to it has the peacemaker cells, in charge of generate the action potential, in other words the natural electric stimulation that was simulated in the left atrium tissue experiment.

Following the same order that have been used, the results are presented, in the first instance for left atria, followed by the right atria results.

5.1 Left atria

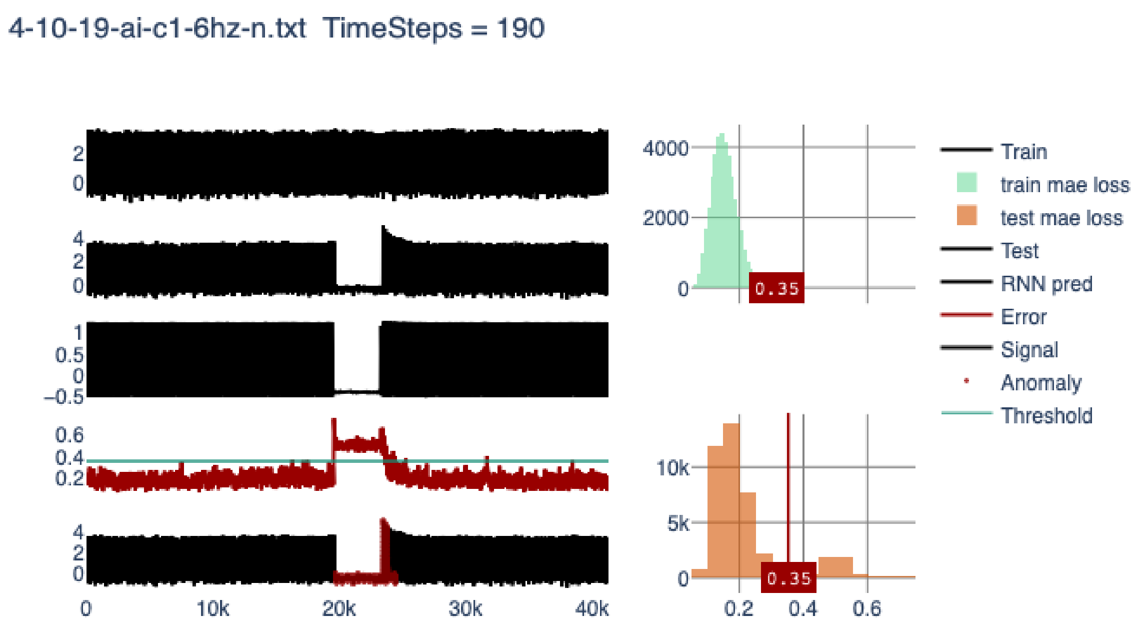


Figure 5.1: 4-10-19-ai-c1-6hz

Evaluating this signal, which was obtained at 4th October 2019, with channel one as objective, it can be seen that the signal with normal behavior has quite same amplitude along the all-time of experiment, the second image, which represents the same one but this time with an empty space at the middle time elapsed, this is one of the anomalies that were mapped in the modeling part, this missing pulses are recognized by the model, as can be observed in the 3rd figure, the threshold for this particular signal is 0.35, in the middle time elapsed the MAE is greater than this threshold for more than $\frac{TIME\_STEPS}{2}$, which is 95 data points, in the windows evaluated, these points are marked as anomalies and are plotted in red color as shown in the last image, this is good detection by the model, considered as success result.

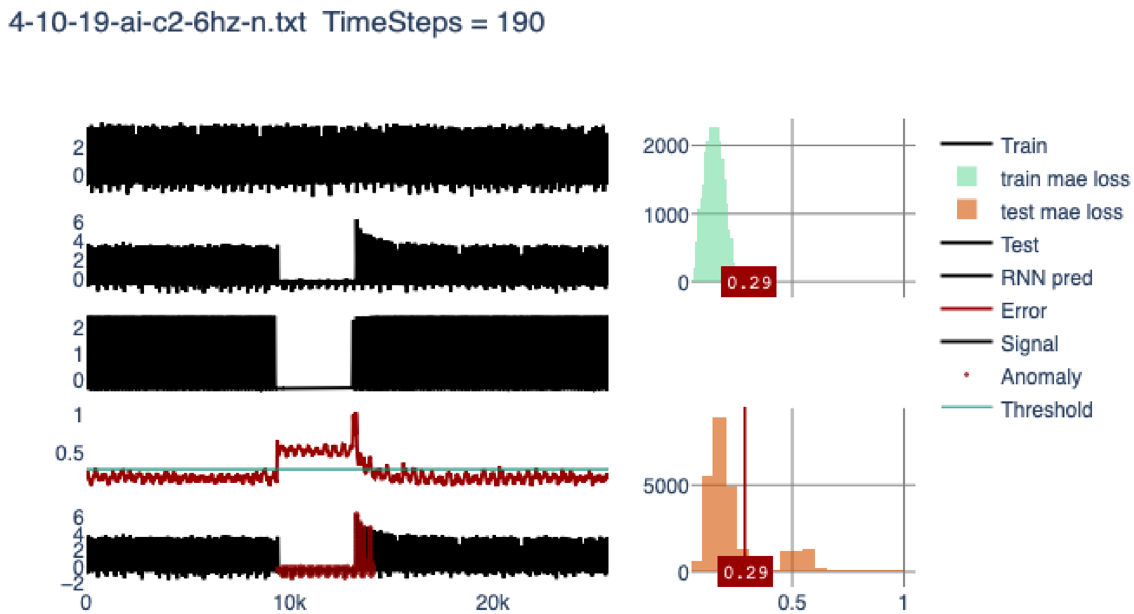


Figure 5.2: 4-10-19-ai-c2-6hz

Actually this signal it's related with previous one, even so the signal is experimented capturing the channel number 2, it's seen that the normal behavior is similar to above image, however the missing pulses last longer, threshold it's greater, giving 0.29, the *MAE* overpass the threshold for more than 95 times in the window chosen, detecting as good as the previous one, and considered as success result too.

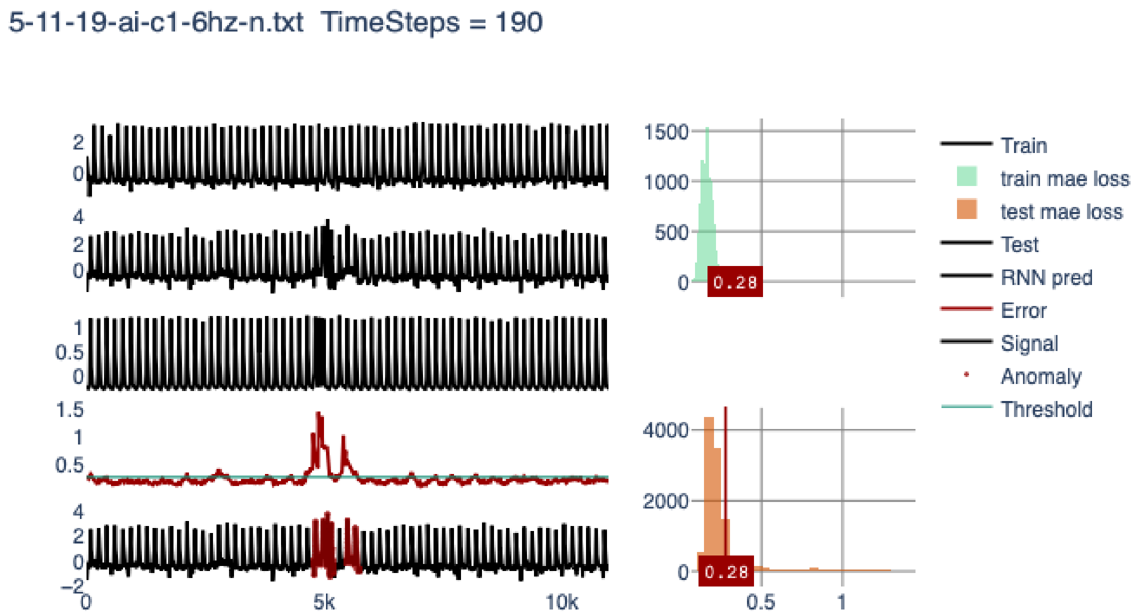


Figure 5.3: 5-11-19-ai-c1-6hz

Current signal was captured at 5th November 2019, with channel one as objective, in the normal signal it can be observed that the pattern is clear, its beats are enough elapsed and are easily visual individualized, on the other hand, the anomalies signal has an erratic pattern in its beats, each one has different amplitudes in the top as in the bottom, although it has two beats consecutively without space in between, when it's observed in the last image, this anomalies pattern is extremely well identified, moreover the network detects other anomalies that with human eye only would be difficult to detect, which are one pattern after and other one before the already explained anomaly, showing the advantage of this approach, this results is considered success too.

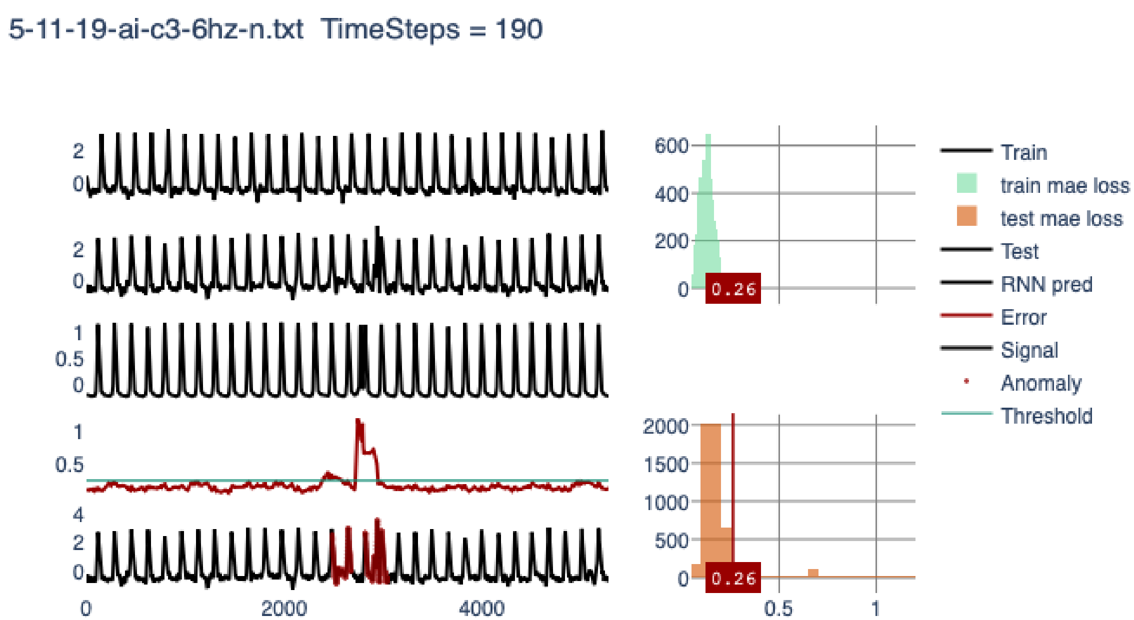


Figure 5.4: 5-11-19-ai-c3-6hz

This signal was captured at the same date, unlike the above one, the channels evaluated was the number 3, the normal pattern showed in the first image, has more elapsed beats than the above one, in the other hand, the test signal has an anomaly at the middle part of the signal, it is called extrasystole, the test signal is quite similar to the training one, unlike the extrasystole, so the anomaly detection detects this pattern and other one before which has a kind of little extrasystole inverted, which is what is expected, it is considered a success result.

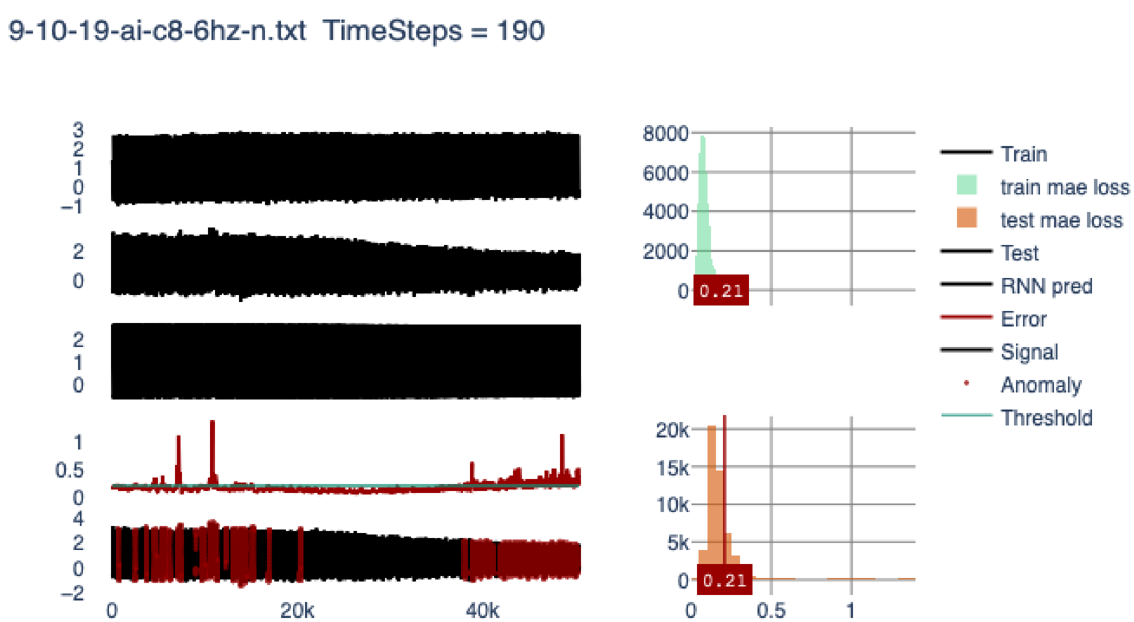


Figure 5.5: 9-10-19-ai-c8-6hz

In this case its evaluated by the model a signal with a static range amplitude as a normal behavior, alternatively the test signal present a descending amplitude from the middle time of experiment to the end, this is called hypoxia event, typically when the cell is getting out of oxygen, its beats' loss strength gradually, moreover at the start of the signal it's not quite similar as training signal, it presents some erratic beats, on both cases the model identify perfectly the abnormal patterns, and points out the reduction of amplitude due to hypoxia and the erratic amplitude at the beginning of the experiment, it is considered a success.

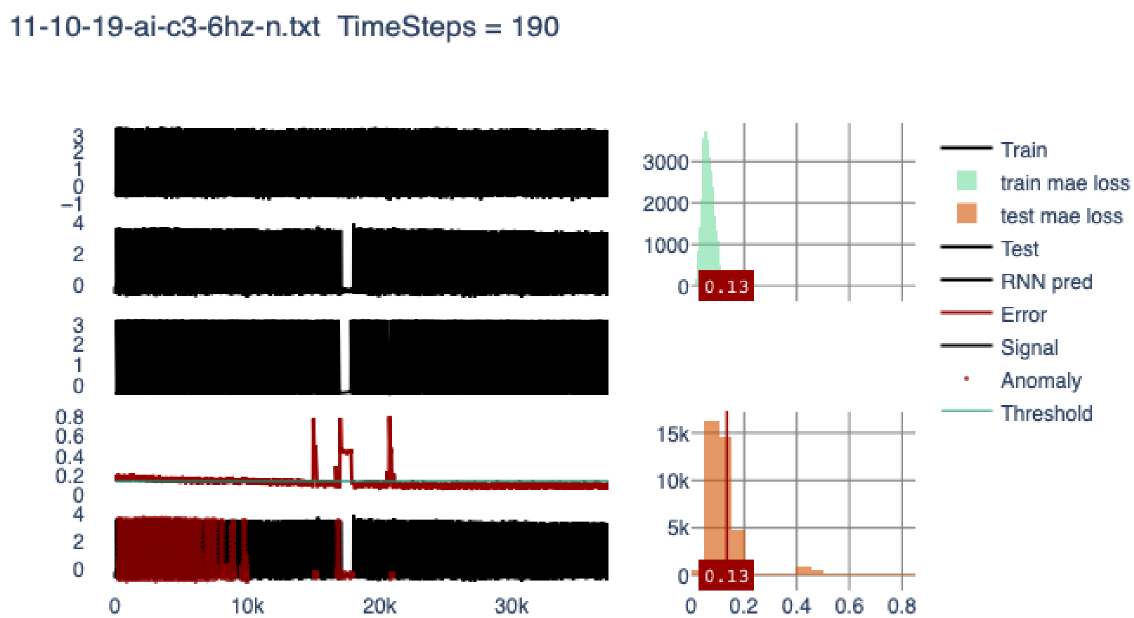


Figure 5.6: 11-10-19-ai-c3-6hz

The present signal represents an experiment done at 11 October 2019, focused on the channel number 3, it's evident the tidy pattern that it present, the test signal, has some beats at the beginning that are larger than other ones, in fact the most evident it's a missing or even a delayed pulse, model prediction is accurate identifying the more amplitude in its initial pulses, missing pulse in the middle and 2 pulses with higher amplitude than the mean, one after and other one before the main missing pulses (white space in middle), network anomaly identification is considered success in this case.

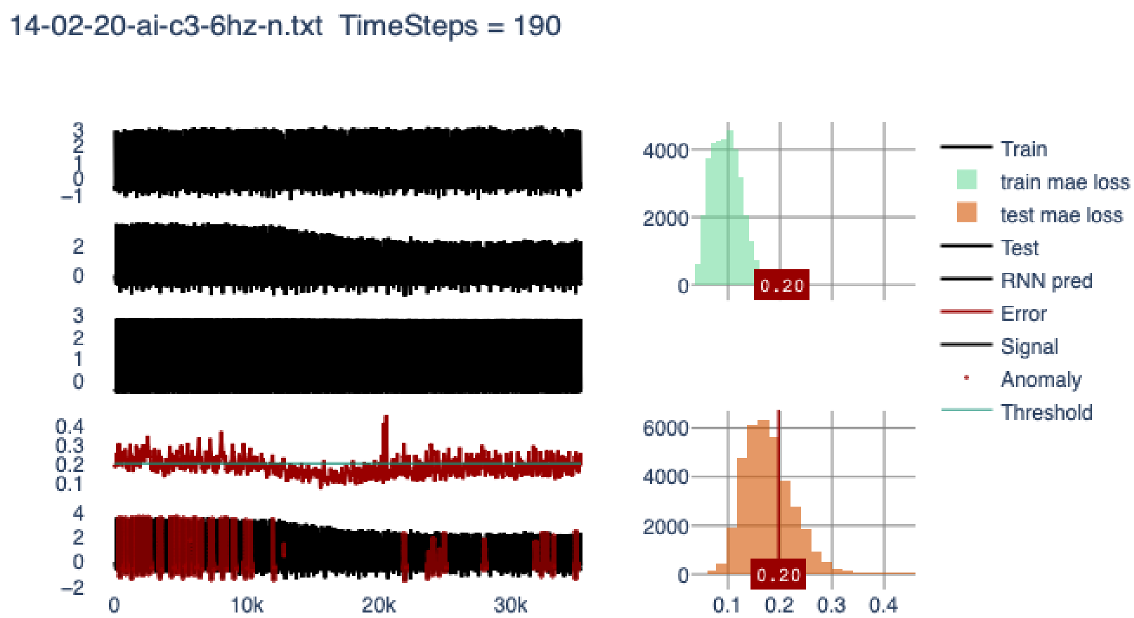


Figure 5.7: 14-02-20-ai-c3-6hz

In this signal obtained at 14 February 2020 with channel 3, it can be observed that follows a constant range over time, but it's a little sloppy at bottom, conversely, the test signal present 2 main possible anomalies that network is expected to detect, one is that the sloppiness of the beats at bottom is higher, second it present a hypoxia pattern, but not as prominent as signal 9-10-19-ai-c8-6hz, taking a closer look of the *MAE* chart, it can be appreciated that at beginning and at the end the *MAE* overpass the threshold, this last one set to 0.20, then at the pulse identification over the test image, it's marked that at the start the erratic pulses and the difference of amplitude it's considered as an anomaly, same for the final part of the signal where the hypoxia reduces even more the amplitude, leaving the middle part, where the *MAE* is under the threshold, as normal behavior compared to training signal at first chart.

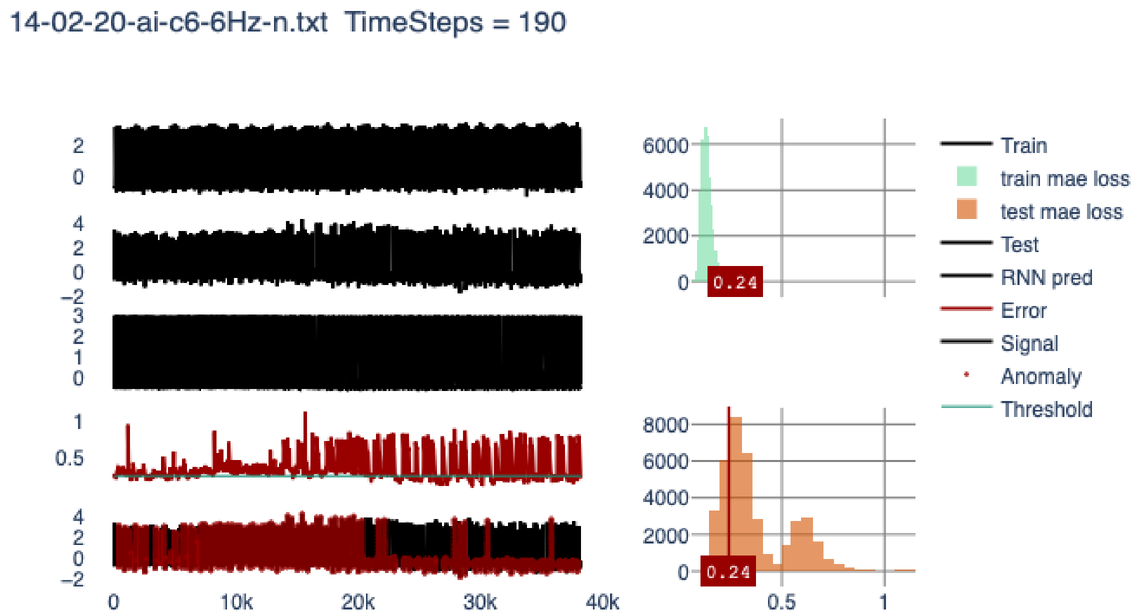


Figure 5.8: 14-02-20-ai-c6-6Hz

Experimental signal from left atrium took the 14 February 2020, with channel numbers 6 as objective, beats goes up and down, in other words present a waving movement, but with constant amplitude in their top as in their valley, by the way the testing signal is not similar as what was explained, presents an irregular waving pattern, some beats in the middle are externally out of normal range, with bigger amplitude than other beats of same signal and even more compared with training data, if the model prediction is analyzed in detail, it can be observed that approximately from 8k pulse, the erratic pattern starts, and it is identified, the outlier beats are also marked, and finally the model detects that the bottom of the rest of the signal presents values that not correspond with the training ones, if the second image is observed back, it has negative values, which training signal does not present, consequently the model identify anomalies correctly and the result is a success.

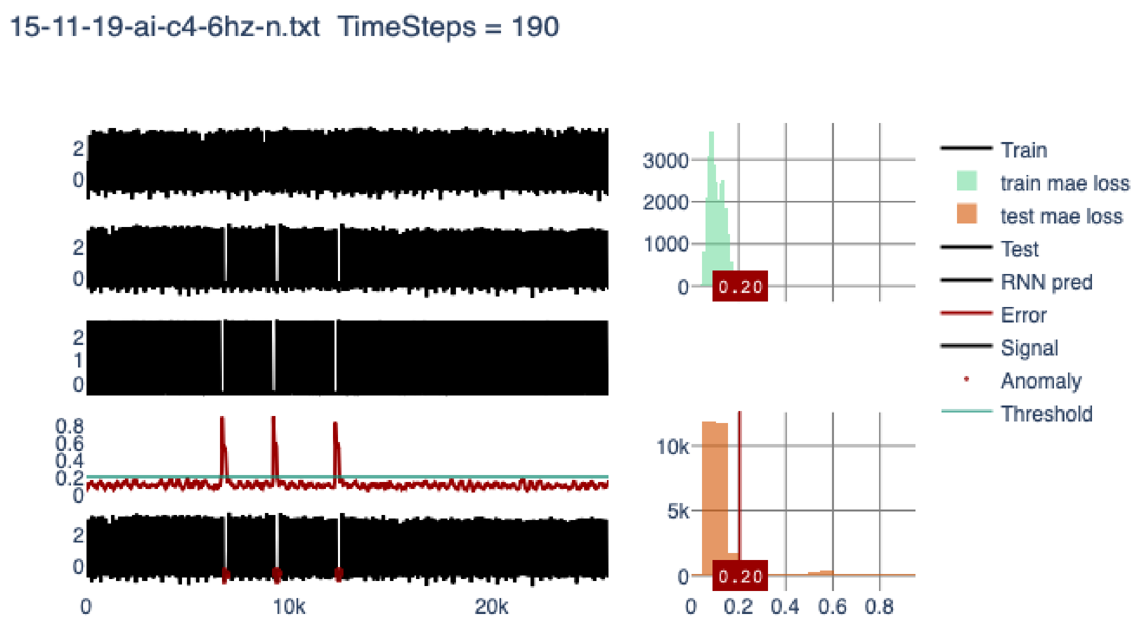


Figure 5.9: 15-11-19-ai-c4-6hz

This signal is one of the best results obtained during developing of this work, what can be seen it is a regular signal with some blur on the bottom and such shorter beats, in the other hand, there is a signal with almost the same range beats, moreover has three missing pulses, in the model recognition these pulses are unmistakable detected, which is an extraordinary result, also for the consistency in many tests, it is marked as a success.

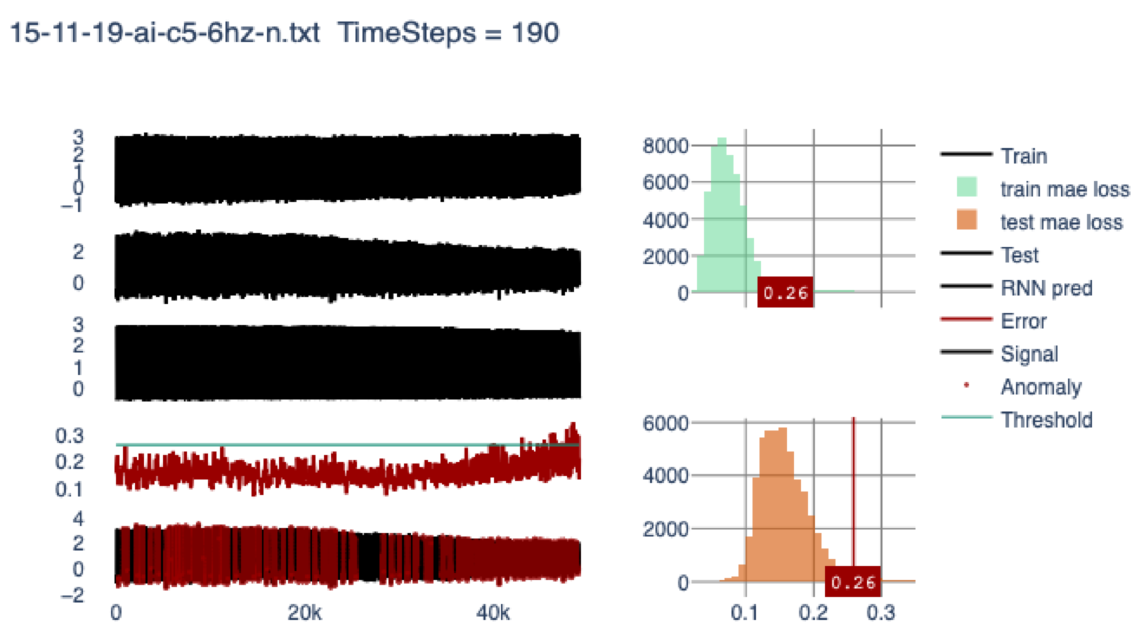


Figure 5.10: 15-11-19-ai-c5-6hz

Finally, signal captured at 15 November 2019 presents a normal signal behavior, with range from ≈ -1 *to* ≈ 3, with an imperceptible decay, on the other side, there's a presence of test signal, which has a lower range in its beats, from ≈ 0 *to* ≈ 2.5, what makes a clear anomaly event a priori, also the signal presents hypoxia pattern from the middle part of the experiment, the model identification result gave the signal

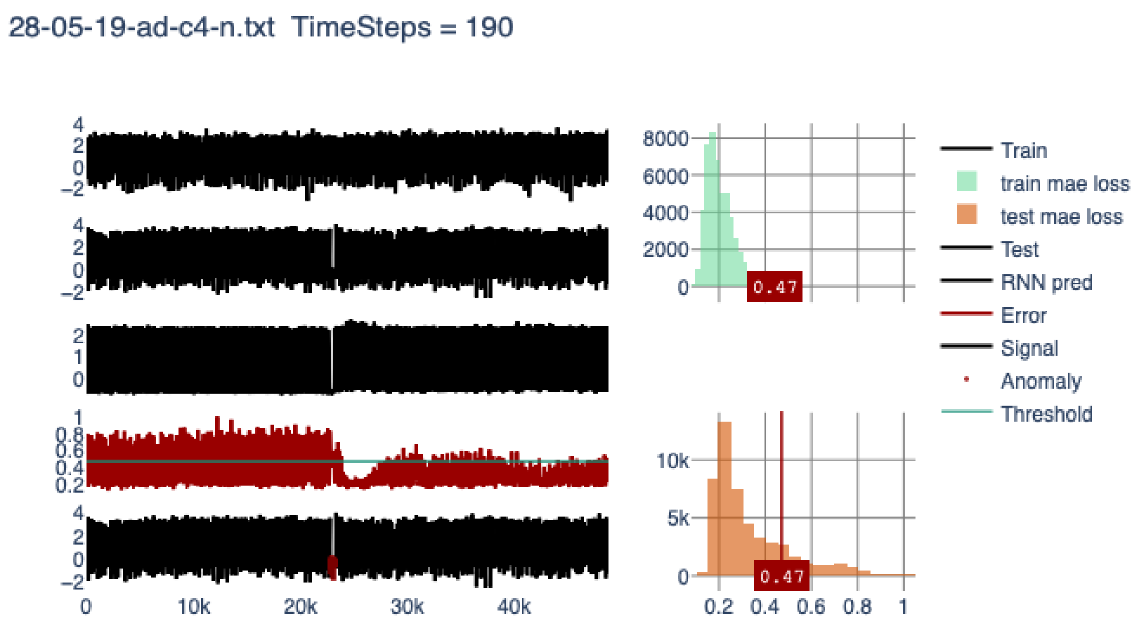


Figure 5.11: 28-05-19-ad-c4

With this current signal captured at 15 November 2019, it can be observed a constant range along the all signal, even though it has an imperceptible upper trend, what makes the range smaller at the end of the experiment, it can be observed also that the range contains negative values, by the way the test signal does not have negative values in its range, some beats seems to be outliers, because they stand out the mean, also the signal suffer a hypoxia event, narrowing the amplitude of the signal in general, the model anomaly identification, shows that the difference in the amplitude, the outliers and the amplitude reduction make up in its entirety an anomaly, only some small parts of the signal are considered as normal events, this test is considered successful.

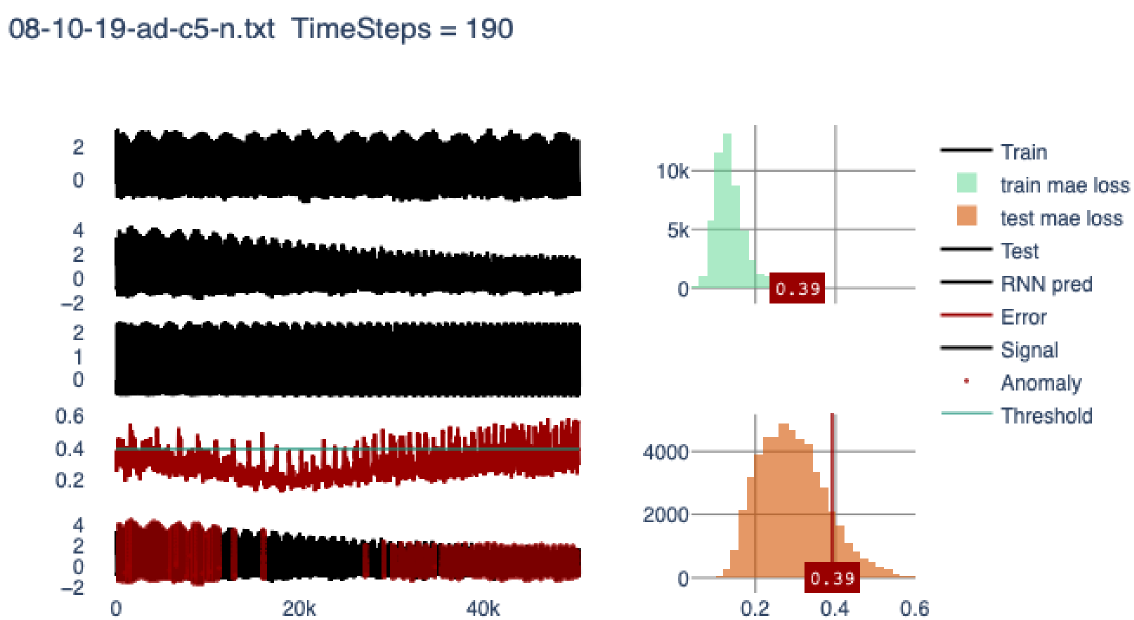


Figure 5.12: 08-10-19-ad-c5

This signal is a good example to explain in detail the window used with the threshold rule, the training signal shows a waving pulses at the bottom of the signal, with some beats larger than usual, negative values around -2. Testing image is quire similar, does no show the waving at the bottom but still the beats larger than usual persist, the anomaly is about the middle of the signal, where a beat is missing, the *MAE* graphic shoes many points overpassing the threshold, but if the last image is observed, it can be appreciated that only the missing pulse is pointed out as anomaly, which is expected, why this happens could be a question, the required number of pulses overpassing the threshold (95) is not accomplished at the beginning of the signal, where seems to be many anomalies based on the *MAE* chart, this rule help to avoid false positives, so this test is considered a success.

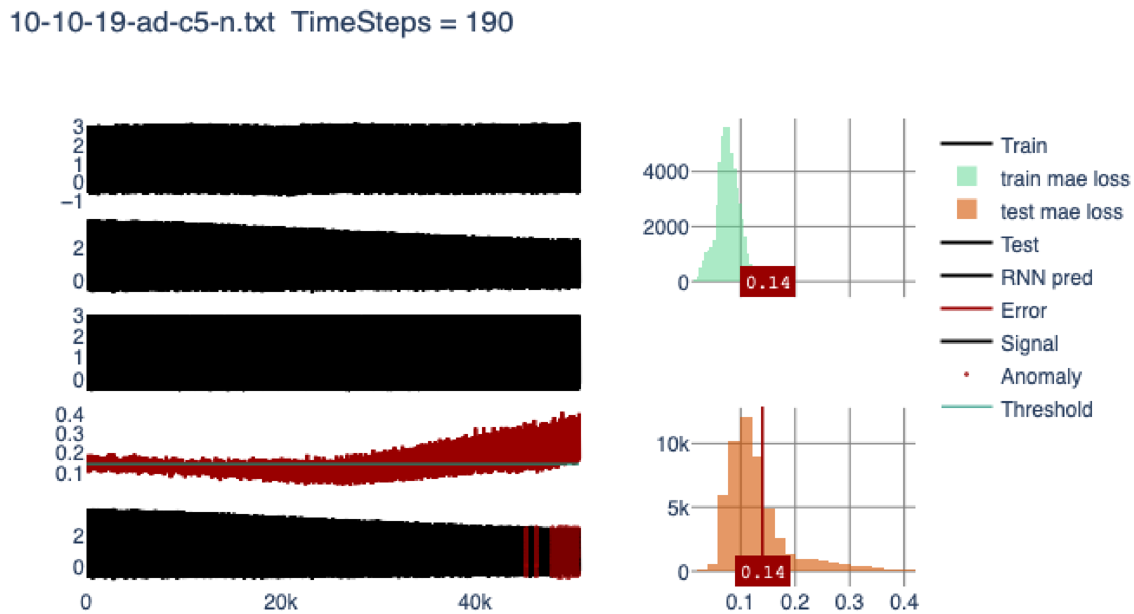


Figure 5.13: 10-10-19-ad-c5

Signal captured at 8 October 2019 presents a marked waving at the top, without any trend and not negative values, testing signal also present waving at the beginning, but it is disappearing gradually, further the hypoxia phenomena is present in high grade, the range in this signal is wider starting from ≈ -2 to ≈ 3, when the last image is analyzed, it shows that at the beginning the range differences are marked as anomalies despite the waving coincidence, at the middle part the ranges are matched, and the network consider that as normal, then the hypoxia events is also marked as anomaly, which is a correct identification, making this test successful

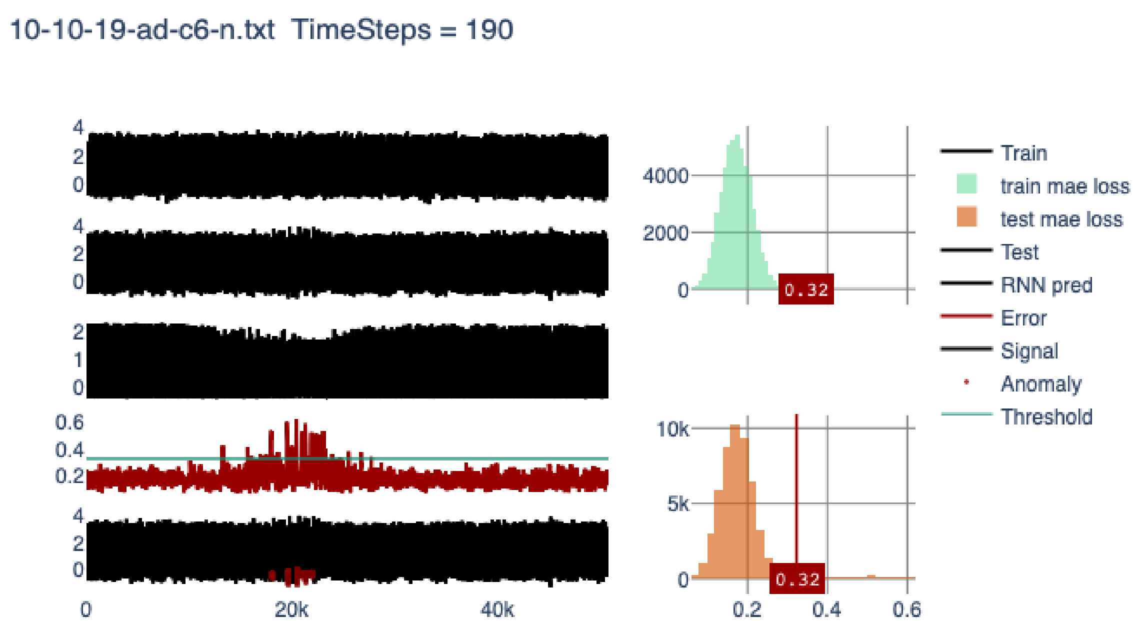


Figure 5.14: 10-10-19-ad-c6

Talking about the experiment done at 10 October 2019, it is considered a clean identification, let's start from the training signal, it has constant range in all the beats and no trend, in the other hand, the testing signal has a constant range in almost all the duration of it, but range goes reducing gradually and little variations, the network prediction detect the last part of the signal as anomaly, because the reduction of the range is enough to make the *MAE* overpass the threshold, which talks about the tolerance of the model, this is considered a complete success too.

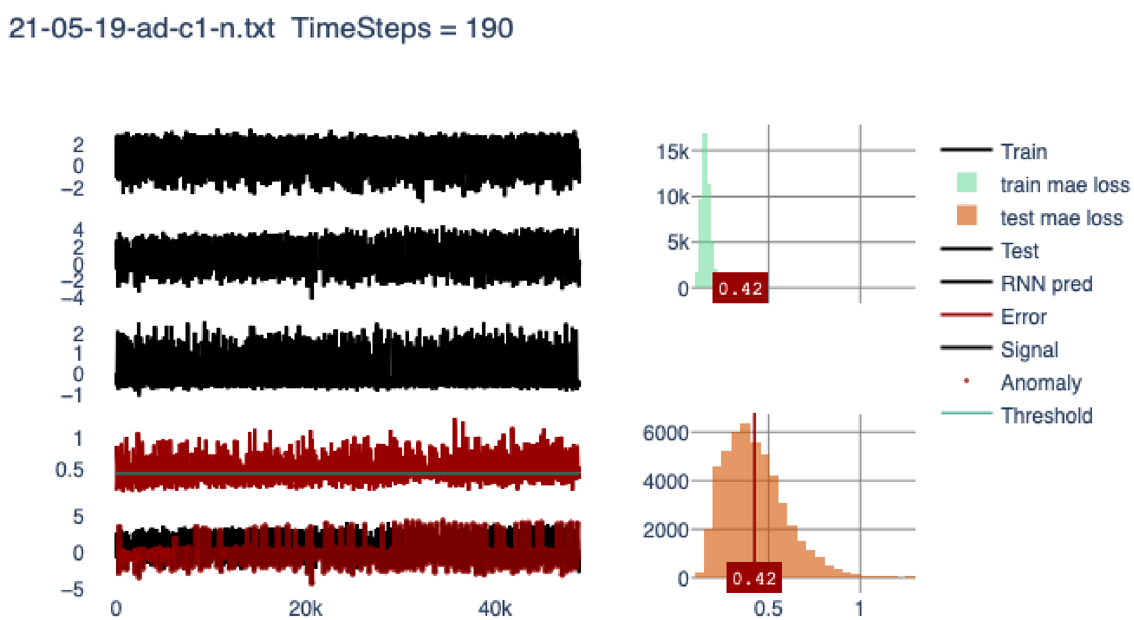


Figure 5.15: 21-05-19-ad-c1

Signal captured at 10 October 2019, in the training signal presents some irregular beats, especially at the bottom, but with consistent range over time, testing image is quite similar despite set of pulses that protrude on tom and other ones in the bottom but more subtle, network apposite of is expected, detected the bottom anomalies and not the top ones, it doesn't mean that the bottom beats protruding are not anomalies, but the top ones are and was not recognized, this is considered a partial success.

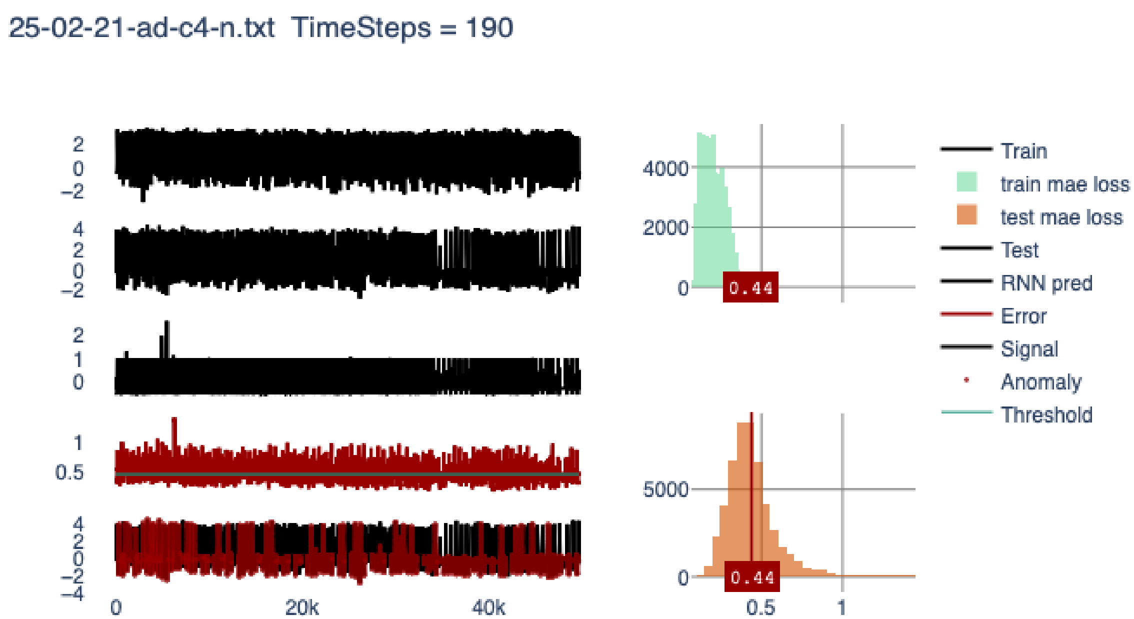


Figure 5.16: 25-02-21-ad-c4

Lastly, a signal with many erratic beats at the bottom in the training signal, a test signal also with erratic pulses and wider range, particularly at the bottom part, the network detect almost all signal as anomaly, which based on the range is correct, but based on the pattern is incorrect, because both signals seems similar in their behavior and in their erraticness, this test is considered a fail, based on the context and pattern recognition that network should show.

6

Conclusion

The purpose of this research was to analyze the behavior of cardiac myocytes signals with external and internal stimulation, and try to identify any kind of anomalies, with help of artificial intelligence, particularly the recurrent neural network, autoencoders and LSTM(Long Short Term Memory) architecture, for accomplish the task, the data processing, problem understanding and method strategy were needed, based on all these activities and results obtained with experimental data, it can be concluded that cardiac myocytes signals can be used to determinate anomalies at cellular level, which will detect earlier a possible heart diseases, allowing early treatment and higher possibilities for patients, this is the main motivation behind this work, having a positive impact in life of patients, all above through researchers and medics work, who are the responsible and experts in the fields, giving them such tools, that allow them to investigate deeper, faster and guarantee quality life for their patients. With data analysis was possible to train a recurrent neural network, with LSTM included, that is capable to learn patterns from normal signals and then receive other one and determinate if it meets the normal (healthy) pattern or not, if not, has the ability to identify the specific location where this is not accomplish and pointed out in red color, indicating what is the abnormal beat or beats, above was proved at the architecture validation with realistic synthetic data 2, later on with experimental data at the results chapter 5, where the model was proven with real data unmodified, from both left and right atrium, containing multiple kinds of anomalies(amplitude, duration, missing beats, etc.), as conclusion of these results, the network performance was extraordinary, from 16 signal samples, 10 from left atrium and 6 from right one, the model had the capacity to detect 14 accurately, the rest two signals, represented cases that must require deeper investigation of how to preprocess that kind of data, which presented a lot of noise and erraticness in their beats; Other great discovery was that the network not only detected the anomalies that were intended to, but also anomalies that are imperceptible to human eye, what shows the potential and the goodness of using artificial intelligence in medical fields, giving relevant information that is difficult to see in the first instance and deducting the false negatives, this last, it is a metric used in artificial intelligence systems, particularly to measure the goodness of the prediction and the confusion of them, predictions have four possibles tags in the confusion matrix, **true positives**, when it is expected that the model predicts it as "true", and it does, **false positives**, when is expected that the model predicts it as “false”, but it predicts a “true”, **true negatives**, when the model is expected that the model predicts it as “false”, and it does, **false negatives**, when the model is expected to predict an event as “true”, but it predicts it as “false”. Each of these tags have a different cost, depending on the problem object of study, when it is treating about people life, the true negatives, are not as costly as false negatives, because a true negative in this work means, that the model predict a pulse as anomalous but actually is normal, what going to happen it’s that the medic would do other test and other procedure and will find out that patient is healthy after all, in the other hand, with a false negative, means that the model predict a pulse as normal when it’s not, so consequently, the medic will let the patient go and will not do other procedure, it will cause that the illness is not detected; Nowadays, the medics performs the model role, making predictions or identifications, results 5 showed that model predicted patterns as anomalies that were not considered as such in principle.

References

[1] (2021). Master in computational engineering and mathematics thesis - results. https://andresospina95.github.io/Thesis_urv_uoc/. Accessed: 2021-06-09.

[2] contributors, W. (2021). Autoencoder. [Online; accessed 27-May-2021].

[3] González-Ormerod, G. (2019). Functional study of atrial contraction in a murine heart model. Master's thesis, Universitat Autònoma de Barcelona.

[4] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith,

N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585, 357–362.

[5] Kingma, D. P. & Ba, J. (2017). Adam: A method for stochastic optimization.

[6] Lees, A. V. (2017). *Multiscale Image Analysis of Calcium Dynamics in Cardiac Myocytes*. PhD thesis, Universitat Politècnica de Catalunya.

[7] McKinney, W. et al. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445 (pp. 51–56).: Austin, TX.

[8] Media, A. M. (2017). Cardiac action potential, animation. https://www.youtube.com/watch?v=v7Q9BrNfIpQ.

[9] Staerk, L., Sherer, J. A., Ko, D., Benjamin, E. J., & Helm, R. H. (2017). *Atrial Fibrillation*. Technical report, American Hearth Association.

[10] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., & SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261–272.