IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 14, 2021, accepted February 26, 2021, date of publication March 9, 2021, date of current version March 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3064854

Semi-Supervised Recurrent Variational Autoencoder Approach for Visual Diagnosis of Atrial Fibrillation

NAHUEL COSTA[1], LUCIANO SÁNCHEZ[1], (Senior Member, IEEE), AND INÉS COUSO[2], (Member, IEEE)

[1]Computer Science Department, University of Oviedo, 33202 Gijón, Spain
[2]Statistics Department, University of Oviedo, 33202 Gijón, Spain

Corresponding author: Nahuel Costa (costanahuel@uniovi.es)

This work was supported in part by the Ministry of Economy, Industry and Competitiveness (Ministerio de Economía, Industria y Competitividad) of Spain/FEDER under Grant TIN2017-84804-R and Grant PID2020-112726-RB.

ABSTRACT In this work we propose a semi-supervised framework to visually assess the progression of time series. To this end, we present a recurrent version of the VAE to exploit the generative properties that lead it to learn in an unsupervised way a continuous compressed representation of the data. We introduce a classifier in the VAE training process to control the regulation of the latent space, allowing the network to learn latent variables that set the basis for creating an explainable evaluation of the data. We use the proposed framework to address the diagnosis of Atrial Fibrillation (AF) first validating it with simulated data with known properties and subsequently testing it with intracardiac data obtained from pacemakers and defibrillator systems.

INDEX TERMS Graphical Analysis, heart disease, recurrent neural networks, time series, variational autoencoder.

I. INTRODUCTION

Most AI-based systems have a black box nature that allows powerful predictions, but cannot be directly explained. This is especially true when it comes to time series data, where the bulk of methods stick to rawly classifying or predicting a number or a set of numbers. Unsupervised learning approaches are a possible alternative for this. Within this paradigm Autoencoders are one of the most promising methods that we can find. Autoencoders are a family of neural networks that have the ability to learn a simplified representation of the data, typically for dimensionality reduction. These networks are designed to reconstruct the input data while at the same time learn a compressed representation of it; the so-called latent space. Variations of the original model [1]–[3] have been developed in order to enhance classification and clustering tasks until the emergence of Variational autoencoders (VAEs), whose main purpose is the generation of new data.

Variational autoencoders are rooted in Bayesian inference [4] and are comprised of an encoder function $q_\phi(z|x)$ and a decoder function $p_\theta(x|z)$ where z is the latent encoding vector, x is the input data and ϕ and θ are parameters that initialize a probability distribution. By introducing the Kullback-Leibler divergence into the loss function, which simply measures how much one probability distribution diverges from another, the above-mentioned parameters corresponding to the input data distribution can be learned. This, together with a reconstruction error added to the loss function, allows the model to produce a latent space in which similar data will be located close to each other and also enables new data to be sampled from points that do not belong to the original data, thus having a generative model.

The main difference between VAEs and the rest of autoencoders lies in the learned latent space: The inputs are not coded to a set of fixed vectors, but the compression depends on a probability distribution $q_\phi(z|x)$ instead, causing the data to be organised in a continuous space, i.e. two nearby points in the latent space should give similar contents when reconstructed (Figure 1). Precisely, other unsupervised techniques such as clustering algorithms lack this property. Although they do prioritize grouping data of a similar nature, the visual disposition of the clusters can often be arbitrary. On the other hand, neither can the VAE latent space be used for clustering

The associate editor coordinating the review of this manuscript and approving it for publication was Giovanni Dimauro.

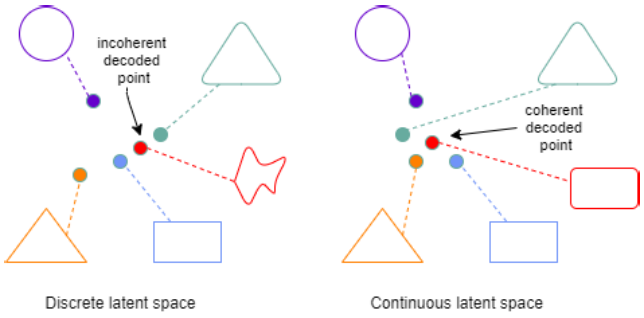

FIGURE 1. Simplified representation of the compression resulting from a vanilla Autoencoder (left) and a VAE (right). When the latent space is continuous, the organisation of the data allows decoding of a meaningful figure, in this case a cross between a rectangle and a triangle, thus favouring the generation of new data.

since the encoded data tend to be overlap to prioritize the generative process.

Therefore, a methodology capable of combining the above properties, that is, depicting the input data into clusters, while preserving a continuous representation according to its underlying complexion, would be of interest to time series data. In fact, there has been a recent interest in seeking such a model, as can be shown in [5]–[8], which make use of Variational Autoencoders together with Gaussian Mixtures in order to achieve an interpretable clustering. Nevertheless, these approaches are not intended to be applied in time series.

Besides, although VAEs have proven to be efficient in multiple domains, mainly related to computer vision [9]–[11] and Natural Language Processing (NLP) [12]–[14], as generative frameworks as well as data compressors, there is a lack of research when it comes to time series. In [15] the authors present a VAE model that can map time series to a latent vector representation, but the model has become obsolete due to more recent advances in recurrent architectures. Other promising work has begun to emerge: In [16] LSTM networks are used to model the temporal complexion of the data, whereas in [17] the authors propose to use echo-state networks for the same objective. Despite the fact that these works combine recurrent architectures with VAEs, their goal differs from ours since they aim to detect anomalies based either on reconstruction errors or on anomaly scores, while what we are pursuing is an interpretable assessment of time series.

The solution that we propose is to introduce a recurrent version of the VAE to deal with temporary data along with the inclusion of a classifier in the training process that controls the regularisation of the latent space to prevent the resulting clusters from overlapping. In this way, a representation that can be used for displaying a graphic map that gives insight into the evolution of the time series is obtained.

The creation of such a model is motivated by the need to offer a solution to a problem in which the presence of efficient algorithms is limited: the diagnosis of Atrial Fibrillation (AF). AF is the most common type of arrhythmia in clinical practice. It is a type of heartbeat in which the atria tremble, causing an irregular and accelerated heart rhythm.

The treatment of the disease often involves the use of pacemakers. These devices are a source of data that record the dates and lengths of the episodes of high atrial rate, comprising a historical record, that is, a time series. Effective and accurate diagnosis of this condition remains challenging these days. Also, a simple prediction may not be informative enough for specialists to examine the state of the disease. Thus, a variational-clustering approach is tailored to our needs in order to accomplish a visual diagnosis capable of assessing the evolution of AF.

The structure of this paper is organised as follows: Section II introduces the importance in the treatment of this condition and the difficulties associated with its diagnosis. A detailed description of the proposed method comprising the semi-supervised VAE framework for achieving an explainable diagnosis is described in Section III. Before reporting experimental results in Section V, an illustrative problem is presented in Section IV while conclusions are drawn in Section VI.

II. AF DIAGNOSIS

AF is an abnormal heartbeat usually presented in the elderly. The course of the disease can lead to a progression from paroxysmal arrhythmia (arrhythmia episodes that appear and disappear spontaneously) to persistent arrhythmia (episodes that last at least seven days and do not end without external intervention) or to permanent arrhythmia (uninterrupted episodes). The progression of AF is a complex process that depends on several risk factors [18], and an early diagnosis may condition the provision of optimal treatment.

Episodes of AF are easily detected on surface electrocardiograms (ECGs), obtained from non-invasive devices, but the activity recorded is over a very specific period of time, which in no case is enough to capture the evolution of the disease. Portable ECG monitors are an advantage in this respect and recent advances in healthcare facilitate continuous monitoring of intracardiac activities. This is beneficial in detecting pathological signatures and arrhythmias [19], especially in patients in the latter stages of permanent AF [20]. On the contrary, the health risks for patients in the early stages of paroxysmal arrhythmia are lower and the disadvantages of wearing these devices continuously outweigh the advantages.

It seems that the situation may change in the near future as new ECG sensors are small enough to be incorporated into wearable devices. The Apple Heart Study [21] shows that different types of AF can be detected in smartwatches, though the battery consumption is high, which prevents the sensor from being always on. To date, the detection and timing of short AF episodes remains an open problem.

In those patients in latter stages of the disease, pacemakers or IDCs (Implantable Cardiac Defibrillators) are normally used to control the heart rate [23] in order to keep common symptoms such as dizziness or chest pain under control.

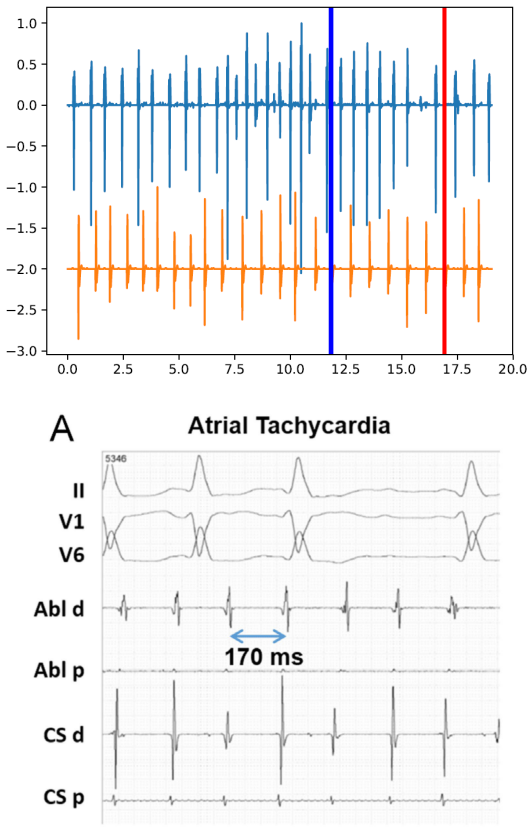

FIGURE 2. Top: Intracardiac ECG. The morphology of the surface ECG is not kept in the iECG, where there is only one peak for each heartbeat. Bottom: Surface ECG (taken from reference [22]).

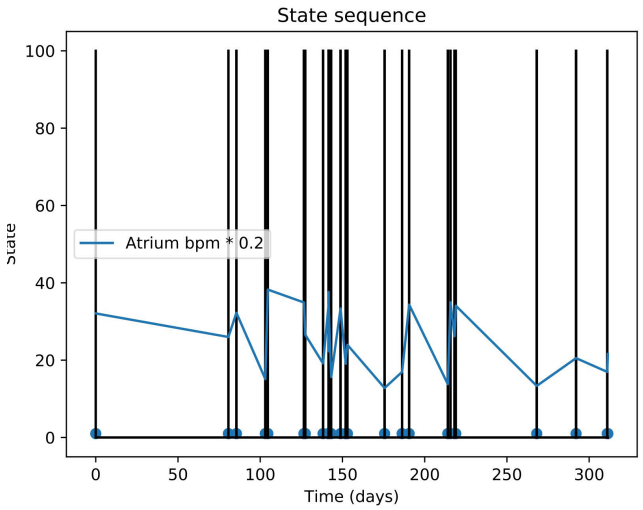

FIGURE 3. Top: Dates of pacemaker mode changes during a year Bottom: Recorded length of the AF episodes.

These devices provide heart rhythm monitoring, being able to detect episodes of arrhythmia, specifically of high atrial rate, which normally correspond to AF episodes. Intra-cardiac electrocardiograms (iECGs) are stored in the memory of these devices (see Figure 2, upper part) which are representations of the difference in potential between two points in contact with the myocardium in space over time. They keep a record of information seconds before and after the detection of each episode. This includes only the instantaneous frequencies of the atrium and ventricle because the morphology of the heartbeat is lost in the high-pass filtering at the IDC electrode, unlike surface ECGs(see Figure 2, lower part). Moreover, this information is not used for diagnosis, but rather to adjust the operational parameters of the device.

Given these facts, the most reliable source of information to work with, are the dates and lengths of the recorded episodes, which are also stored. This is a drawback because patients with multiple episodes will probably be in the latest stages of AF, but the interest lies in patients with a short history (initial episodes) so that the worsening rhythm of AF can be anticipated. This goal is very challenging because it is difficult to find a model that fits such a small amount of data (see Figure 3). Besides, the fact that the data are non-stationary makes the problem even more complicated, and this is exactly what we want to predict on the basis of a short sample, the transition from paroxysmal to permanent arrhythmias.

There are additional difficulties [24] because the algorithm used by the IDC to determine the duration of the episodes is not completely reliable. The device parameters are adjusted based on the iECGs mentioned above and safety concerns prevail, so the false positive rate is high. This leads to long episodes of AF that are sometimes mistakenly reported as short episode sets, so preprocessing is needed to take these spurious events into account, which in turn causes the number of episodes to be reduced even further.

All these reasons being explained, it is clear that the progression of AF is a complex process that depends on many different factors. To ease the interpretation of these factors, we are looking for a model capable of providing clinical staff with a diagnostic tool that can accurately determine the status of a patient with AF, therefore something more than a straightforward prediction is pursued. A possible path is to establish a Representation Learning approach since, unlike others, the performance of models following this approach depends directly on the internal representations, which in turn can be leveraged in favour of a better understanding of the problem itself. Typically, an algorithm capable of learning the characteristics that best represent the underlying data distribution is required, making it easier to perform other tasks such as classification or prediction. Since Principal Component Analysis (PCA) was developed, Representation Learning has been investigated to overcome the challenges of high dimensionality. Over the last decade, Deep Learning has been taking an important role in this field through supervised and unsupervised learning strategies, where it has had a great impact due to the feasibility of processing temporal/spatial data or images more efficiently than superficial methods such as ICA, LDA or LLE.

Representation Learning has been employed in several areas of medicine for purposes such as risk factor selection,

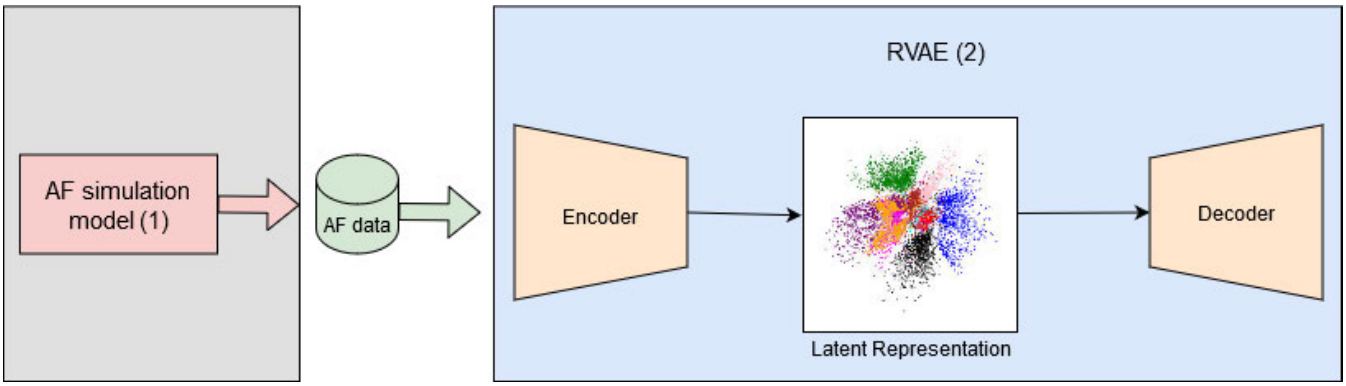

FIGURE 4. Pipeline of the proposed solution. A simulation model is used to generate synthetic arrhythmias that reflect different stages of Atrial Fibrillation. These data are fed to the proposed recurrent VAE so that it learns a representation that will later be used to evaluate the condition of new patients.

disease phenotyping, and prediction or classification of disease risks [25]. This line of research is key to developing explainable AI, where the results can be interpreted by human experts. There is a lack of work in this direction concerning the diagnosis of cardiovascular diseases where the few works that exist are focused on image processing [26] or simple classifiers are developed for time series data [27]. Regarding the diagnosis of AF, the vast majority of papers analyze ECG data from non-invasive devices [28], which are compatible with patients in the early stages of the disease or without previous pathologies, hence they are out of the scope of this work. There are also incentive contributions [29] where the authors study data from wrist-worn devices with convolutional networks. Nonetheless, pacemakers are still the devices that can provide valuable information in those patients in more advanced stages of the disease.

In our previous work [30] we tried to contribute to this path by presenting a graphical approach for analyzing the progression of AF using the output of Recurrent Neural Networks (RNN). Activations of the last layers of LSTM and GRU classifiers were used to create a topological map that provided an intuitive visual diagnosis. Although the results achieved were significant in terms of accuracy, the interpolation used to create the map obtained was highly sensitive to the differences between the neuron activations, which might provoke inconsistencies in the map. For that reason, a recurrent version of Generative Adversarial Networks (GAN) [31] was also introduced to use an ensemble formed by the discriminative part of these nets trained on different types of arrhythmia in order to learn a representation according to the complexion of the data. Nevertheless, the classification results were outperformed by the LSTM and GRU classifiers.

What we propose in this work is to exploit the representation learning potential of VAEs in order to provide a visual early diagnosis of the evolution of AF. VAEs have the ability to condense data from a high dimensionality to a much smaller dimension, maintaining consistency between the distance of the data that is reduced. The influence of Bayesian Variational Inference provokes the learned latent space to depict a two-dimensional projection of the data according to its nature, which is accurate to the AF problem: this resulting latent space can be interpreted as an explainable map where the distances between the compressed input data, are correlated with the differences between the various types of arrhythmia. When a sample of a patient's data is presented to the VAE, the resulting location on the map gives insight into the state of the intracardiac activity of the patient and how the disease might evolve in a short period of time, due to proximity to other nearby points.

III. PROPOSED METHOD

This section describes the proposed framework for performing the task of evaluating the evolution of time series. Figure 4 shows the pipeline followed for applying this framework to the diagnosis of patients with AF out of their intracardiac data, where two main components are distinguished: First, a model capable of simulating the behaviour of actual AF clinical data (1) is used to generate a dataset that reflects the variety of arrhythmias that a patient may suffer. Then, a recurrent VAE (2) is trained with the generated dataset and consequently, a latent representation that serves as a basis for creating the proposed diagnostic tool is obtained.

A. AF SIMULATION MODEL

One of the main difficulties in applying Machine Learning methods to medical problems is the data availability. It is well known that the larger and more diverse the dataset with which the model is trained, the better the learning. However, medical data are often highly sensitive and there are privacy concerns. In this case, although pacemakers registers can be collected keeping the privacy of the patients, gathering enough data that reflect the different progressions the disease may have in different people is beyond our reach. Instead, we opt to use the simulation model presented in [30].

This simulation model is based on a continuous Markov model with 3 states: ''Normal'', ''Arrhythmia'' and ''False Normal'' (see Figure 5). The first and second states refer to the periods of time in which the patient is in a normal

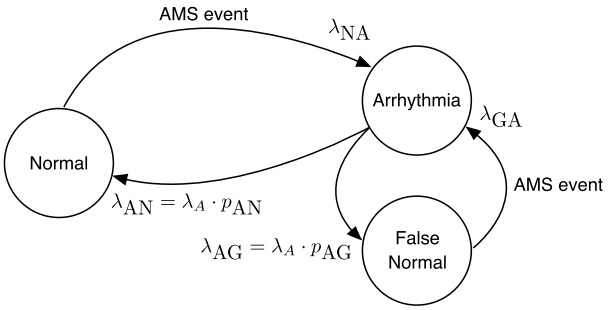

FIGURE 5. State diagram of the dynamical model of the simulation of AF episodes.

state (out of arrhythmia) or suffering an episode. The ''False Normal'' state refers to those cases, as stated in the previous section, in which the pacemaker erroneously detects the end of an episode of AF and which subsequently leads to a change in the pacemaker's operation mode to control the arrhythmia, a process also known as Automatic Mode Switching (AMS). There are AMS events in the transitions from ''Normal'' to ''Arrhythmia'', but also in the transitions from ''False Normal'' to Arrhythmia.

AMS events can therefore define the beginning of a true episode of AF or the false end of an episode (''False Normal'' state). The latter case is not desired but there is no simple procedure to purge these events from the IDC data in real patients [32], so the proposed generative model should produce these spurious events as well.

It will be assumed that the times in the ''Normal'' and ''False Normal'' states follow an exponential distribution with parameters $\lambda_{NA}(t)$ and $\lambda_{GA}(t)$, respectively. The time in the ''Arrhythmia'' state follows an exponential distribution as well, with parameter $\lambda_A(t)$. The probability that the next state after ''Arrhythmia'' is ''False Normal'', where the end of an episode is signaled before time, is p_{AG} and the probability that instead of ''False Normal'' it is ''Normal'' is $p_{AN} = 1 - p_{AG}$.

Under these conditions, the parameter $\lambda_{NA}(t)$ determines the distribution of the time between two episodes and the parameter $\lambda_A(t)$ determines the duration of an episode. The progression from paroxysmal to permanent AF is measured by the rate of change in these two parameters: the time between episodes will be shorter and their duration longer as the heart condition worsens. The rate of progression is modeled by a parameter $\alpha \in [0, 1]$,

$$\lambda_{NA}(t) = \lambda_{NA}(0) \cdot \alpha^t, \quad (1)$$

$$\lambda_A(t) = \lambda_A(0) \cdot \alpha^{-t}, \quad (2)$$

where $\alpha = 1$ denotes a stable patient while α values less than 1 evidence patients with a quick progression to permanent arrhythmia.

To sum up, the proposed generative model is a Markov model in continuous time characterized by 5 parameters: $(\lambda_{NA}(0), \lambda_{GA}, \lambda_A(0), p_{AG}, \alpha)$. With this model, it is possible to produce a list of AMS events through Monte-Carlo simulation by using a random seed. Each randomly generated list can be viewed as a hypothetical patient whose type of AF is defined by the above parameters. The data generated by this model will be used to create the training set for the proposed VAE.

B. RECURRENT VAE (RVAE)

The workflow followed in this component is quite simple: a VAE is trained with the generated dataset to learn a simplified representation of the data. Thus, the learned encoder acts as a feature extractor that describes the input data according to its properties, which are different stages of AF. This section explains how this extraction, reflected in the resulting latent space, can be leveraged to create the diagnostic map we are pursuing. It also emphasizes the recurrent architecture proposed to deal with time series as well as how the presence of a classifier built over the frozen weights of the encoder in the training process can influence the final solution.

1) ENCODER AS A FEATURE EXTRACTOR

In a VAE the training is regularised to avoid overfitting and to ensure that the latent space has good properties that allow the generative process. Precisely these properties contribute to the input data being mapped in the latent space in such a way that similar data are nearby and that this representation can be used as a feature extractor.

A VAE, given an input, tries to find a latent vector that is capable of describing it and at the same time has the instructions to generate it again. The process can be described as: $p(x) = \int p(x|z)p(z)dz$. Given that the integral of this formula is intractable due to the continuous domain of z, the variational inference is needed via the lower bound of the log likelihood, $\mathcal{L}_{vae}$,

$$\mathcal{L}_{vae} = E_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - D_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{x})||p_\theta(\mathbf{z})). \quad (3)$$

The first term is the reconstruction of x that tends to make the coding-decoding scheme as efficient as possible by maximizing the log-likelihood $\log p_\theta(\mathbf{x}|\mathbf{z})$ with sampling from $q_\phi(\mathbf{z}|\mathbf{x})$, modeled by a neural network whose output are the parameters of a multivariate Gaussian: a mean and a diagonal covariance matrix. The second term tends to regularise the organisation of latent space by causing the distributions returned by the encoder to approach a standard normal. It regularises the latent variables (represented by z) by minimising the KL divergence between the variational approximation and the prior distribution of z. The encoder, represented by $q_\phi(\mathbf{z}|\mathbf{x})$ is the component that will be used as a feature extractor since its goal is to map the input data into a lower dimensional space.

One of the advantages Autoencoders have is the flexibility provided by their architecture. The temporal nature of the data suggests that the VAE can be combined with time series modeling approaches such as RNNs. Among the different types of RNN that can be found, LSTM networks are the most outstanding ones. LSTM networks process data from back to front preserving the information from the past through the hidden states. Nevertheless, it is also possible to preserve

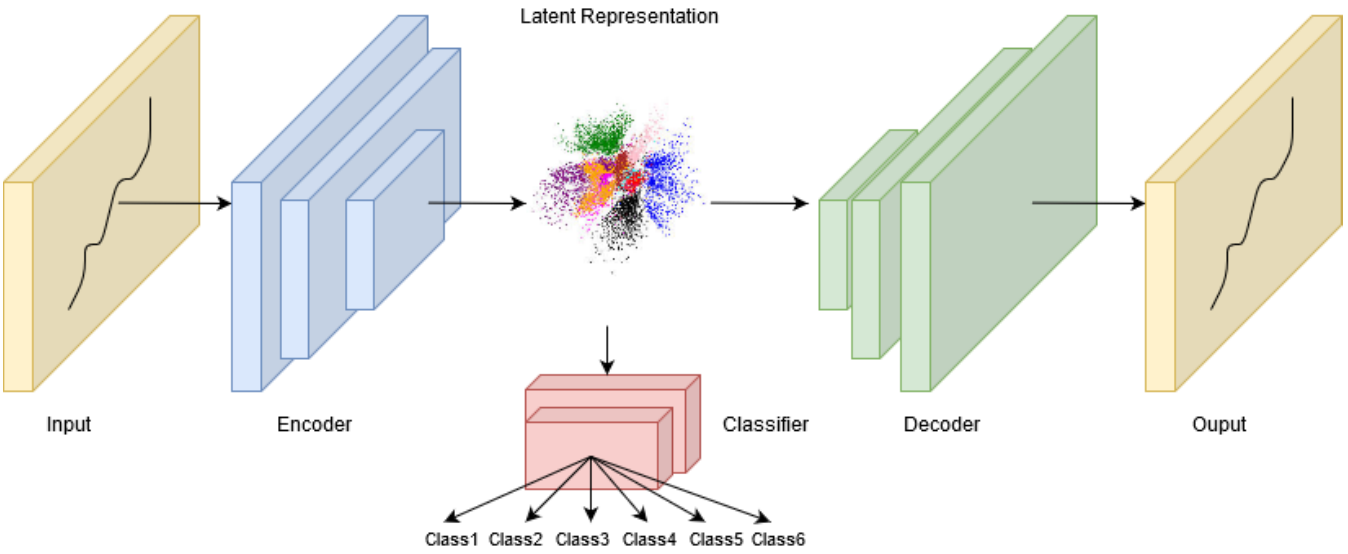

FIGURE 6. Network structure of the proposed method. The blue and green blocks are the encoder and decoder respectively and the red blocks refer to the linear classifier.

information from the future by processing data from front to back. This property is the operating principle of Bidirectional LSTMs: they run in two directions, from past to future first, and then from future to past by preserving information from both periods. This is very valuable due to the fact that the network is aware of how the data may look like in its future stages, so it can help to understand what kind of information to predict (different progressions of AF).

With that being said, we decide to replace the encoder of a vanilla VAE with a Bidirectional LSTM network. In this way, the encoder approximates the Gaussian distribution $p_\theta(\mathbf{z})$ by feeding the output into two linear modules to estimate its mean and covariance. The compression of the input data results in a two-dimensional latent space dominated by the axis represented by the mean and the variance of the approximated distribution. It is expected that arrhythmias are grouped in different clusters according to their features, depicting a simpler representation of their nature.

Based on the representation learned by the encoder, the data, x, is sampled from the conditional probability distribution p(x|z). For generative purposes, this regularisation in the latent space is very effective for easy random sampling and interpolation for the creation of new data. This is the objective of the decoder and is the most extended application of VAEs in the literature. Yet, we decide to discard this part after training the model because our efforts are focused on the diagnosis of the input data instead of the generation of new unseen cases.

2) DIAGNOSTIC MAP

The diagnostic tool introduced in this study is a color-coded map that displays the actual state of the patient and the speed of change in his/her condition from paroxysmal to permanent AF. Once the VAE is trained with the Monte-Carlo simulation of pacemaker events, a topological map is obtained in the latent space from which evident clusters corresponding to different types of arrhythmias are identified, as can be seen in Figure 7 (part right). In this respect, the map can be regarded as a graphic projection of the pacemaker data in a space whose coordinates are the values of λ_{NA}, λ_{GA}, λ_A, p_{AG} and α. The values of λ_{NA}, λ_A and α in the projection measure the condition of the patient and the progression of the AF. λ_{GA} and p_{AG} measure the chance that an AMS event in the pacemaker is spurious. When actual pacemaker registers are used as input, the encoder will place them according to their features, giving information about what type of arrhythmias the patient suffers depending on which cluster they fall into. The following section provides further details on the interpretation of this map.

3) CLASSIFIER: ENCODER QUANTITATIVE DIAGNOSIS

In order to have a clustering-like approach from the latent representation that the encoder learns we propose the inclusion of a classifier in the model training process. A similar approach was taken in [33] by Kingma *et al.* in what they refer to as the latent-feature discriminative model. The authors train a VAE and then feed a classifier with the outputs obtained from the resulting encoder. That is to say, a classifier is used to enhance the benefits of the VAE; however, what we propose is to enhance the latent clustering properties of VAE by using a classifier, not once the VAE is trained, but while it is being trained.

The image on the left in Figure 7 represents why the first approach is not suitable for the problem at hand. It corresponds to the latent space the encoder learns after training the model without any restrictions, therefore the regularisation of the latent space for the generation of new data is prioritized. This results in the location of the input data in areas where

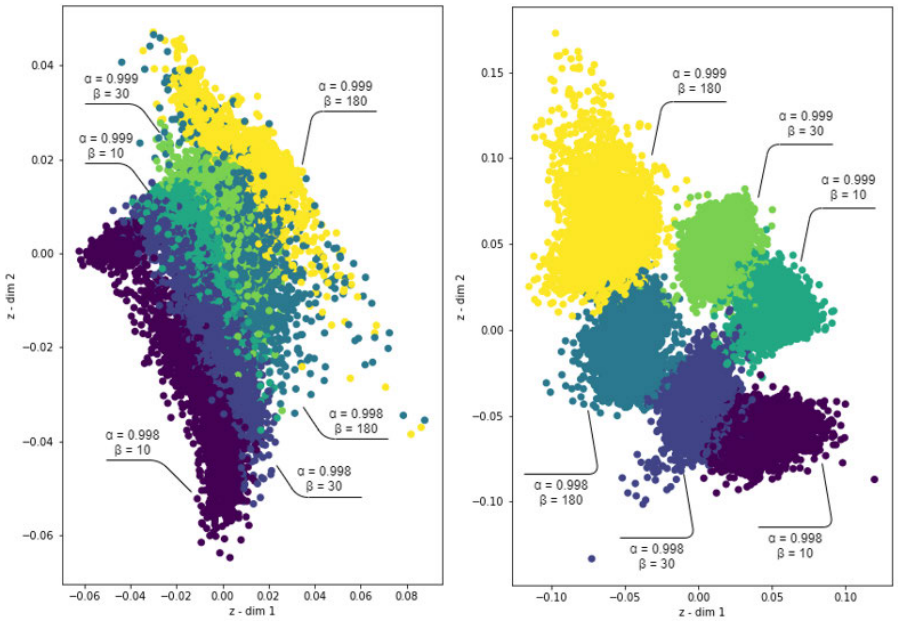

FIGURE 7. Latent representation learned by the encoder following two different training approaches. In the left figure no restriction was added to the model while in the right one a penalty for misclassifications was included.

instances that do not belong to the same group of parameters with which they were trained, are located nearby and in many cases overlap. A classifier was trained over the frozen weights of the encoder, nevertheless, this overlap severely penalizes the performance in classification.

Instead, we decided to take another path: including the optimization of the classifier in the training process. Although few previous works have taken this approach the results are very promising [34]. Therefore, unlike what is proposed in other works applied to VAEs where the problem is divided into two steps: an unsupervised pre-training step (VAE), followed by a supervised learning step (classifier), we decide to merge both. The VAE is trained to minimise a loss function composed of two objectives:

$$\mathcal{L}_{vae} + L_{cross-entropy}(y, \widehat{y}) \tag{4}$$

The first objective corresponds to the VAE objective itself and the second one is the categorical cross-entropy between the labels and predictions for measuring the performance of the classifier. By including the optimization of the classifier in the loss function a restriction is added to the VAE because it will strive not only for a continuous latent space but also for a space where the different classes are separated enough to be clearly differentiated, as can be seen in the right side of Figure 7. The architecture of the classifier is simple: A single Fully Connected layer and a softmax layer are added on top of the encoder base.

In addition to the visual diagnosis that can be offered in the map explained in the previous subsection, the classifier obtained can report explicitly which parameters modeled by the simulation model are the ones that best represent each arrhythmia that is fed to the model. Also, the fact of building a classifier will allow us to compare our method with other state-of-the-art classifiers as we will see in the following section.

Coming up with this result was not straightforward. Firstly, a good choice of the Learning Rate (LR) is necessary, and secondly, it must be taken into account that two optimizations are being made: the VAE objectives and the classification objective. This means that the contributions of each loss must be assessed. Nevertheless, the relevance of each objective is unbalanced and this causes the model representations to preferentially optimize the task with the highest individual loss. To solve this, we decided to use a penalty for misclassification by using different weights for each problem, thus, we can find the perfect balance for the objectives we pursue.

IV. ILLUSTRATIVE PROBLEM

Before addressing the diagnosis of arrhythmias, we present a generic problem to demonstrate the ability of our model. We aim to develop a solution that can analyze sequential data by presenting a visual interpretation of its nature. To this end, a dataset composed of sinusoidal sequences will be used as an illustrative example of what can be obtained with our framework. Thus, we have a dataset whose nature has a periodic factor and relies on three parameters: frequency, amplitude and phase. We generate six classes by varying the frequencies in [1.0, 4.0], amplitudes in [0.125, 0.5], and random phases between $[-\pi, \pi]$.

The aim of training the model with these data is to obtain a representation that is capable of dividing the six classes into different groups and at the same time keeping a coherence

between the distances of the different clusters. That is, if a class has frequency 4 and amplitude 0.125, it is not desired that samples belonging to this class are grouped near those belonging to the class generated with frequency 1 and amplitude 0.5 because the dissimilarities are evident and what we pursue is that data be located near those that are most similar.

Figure 8 (part right) shows the resulting latent space after training the model. At first sight, there are six clusters, each one belonging to each generated class. In terms of classification, the performance is optimal since each point is classified within the group to which it belongs and the proximity between clusters, which is the feature we will use later for the diagnosis, is understandable. The 3 most external clusters belong to examples that have the same amplitude: 0.5, but different frequency: 1, 2 and 3 from most external to most internal. The 3 innermost clusters have the same frequency: 4 and 0.125, 0.25 and 0.375 as amplitudes from innermost to outermost. This arrangement shows that the most similar classes are adjacent on the map, which may let a new point be located in the area of the map that best fits its parameters. Again, it is important to highlight the influence of Bayesian Inference in the display of the clusters since other clustering methods lack this property as can be seen in the left side of the figure where PCA was used. Sequences with frequency 4 and amplitude 0.125 and 0.25 are not clearly differentiated and most importantly, the outermost cluster belongs to sequences with amplitude 0.375 and frequency 4, however, according to the similarity between the data it should be placed between the blue and red clusters. In consequence, the importance of achieving a faithful representation according to the similarity between the data is appreciated.

To conclude this section, it should be noted that although the reconstruction is remarkable (see Figure 9), the addition of the classifier provokes some division between clusters. This penalizes the generative condition because if sequences are generated from ''empty'' latent zones, that is, where there are no points previously represented, the resulting reconstruction may not make sense. Nevertheless, the generative purpose of the framework does not fall within our objectives nor will it be used for any purpose.

V. EXPERIMENTS

The experimental validation of the proposed framework has two parts. First, synthetic data with known properties are used to compare our framework with other state-of-the-art classifiers. Second, actual patients are diagnosed and their maps are validated by a human expert.

We begin by describing the experimental setup and then introducing the numerical results. Finally, the diagnostic map achieved is presented and the experimental validation is discussed.

A. EXPERIMENTAL SETUP

In the experiments carried out for both, the toy problem and the AF diagnosis, the datasets were composed of sequences of length 144 and one feature. The number of samples for training was 84000 and 16800 for test, which were completely balanced between the six classes that were used. On the other hand, as hyperparameter tuning is a very challenging task, we made use of Hyperopt [35], a specific library for hyperparameter optimization. Also, an accurate choice of the LR is particularly essential to improve the optimization process, therefore for this parameter we used an adaptive LR optimizer, Adam, and the Cyclic Learning Rate technique proposed in [36] to help to select the optimal LR with which to start the training. Also, to attain the best possible performance of our model we used callbacks in the Keras [37] Deep Learning library for all our experiments to relegate the training stop condition to the validation error instead of the number of epochs. These implementations led us to find the best results in terms of the functions to be optimized. All models and experiments were implemented in python and the source code to reproduce the experimental results is available in a public git repository: https://github.com/NahuelCostaCortez/RVAE.

B. NUMERICAL RESULTS

In this section, we demonstrate that our framework can compete with state-of-the-art classifiers for time series for the case at hand. It should be noted that the baseline methods we present do not include any representation of the data, but simply predict the class to which each sample belongs, which makes us appreciate the importance of Representation Learning as it provides a more illustrative information than just a numerical or categorical result, as we will see in the next subsection.

In regard to the data, six AF categories were generated using the model described in Section III. These classes are labelled `998na10`, `998na30`, `998na180`, `999na10`, `999na30` and `999na180`. The class labels begin with the first three decimals of α, which is the speed of the progression of the AF (998 is fast, 999 is slow) while the second number in the class label is $1/\lambda_{NA}(0)$, from now β for simplicity, which is the average time between two AF episodes, measured in days (10, 30 and 180 days).

1) BASELINE METHODS

To evaluate the performance of our model, we used the tool provided in [38] to implement 5 baseline methods,

- Resnet: a deep Residual Network proposed by [39] composed of three residual blocks followed by a GAP layer and a final softmax classifier whose number of neurons is equal to the number of classes in a dataset.
- FCN: A Fully Convolutional Neural Network, with the architecture also proposed in [39], which consists of three convolutional blocks whose result is averaged over the entire time dimension that corresponds to the GAP layer. Finally, a traditional softmax classifier is completely connected to the output of the GAP layer.
- Encoder: Originally proposed by [40], Encoder is a hybrid deep CNN whose architecture is inspired by FCN

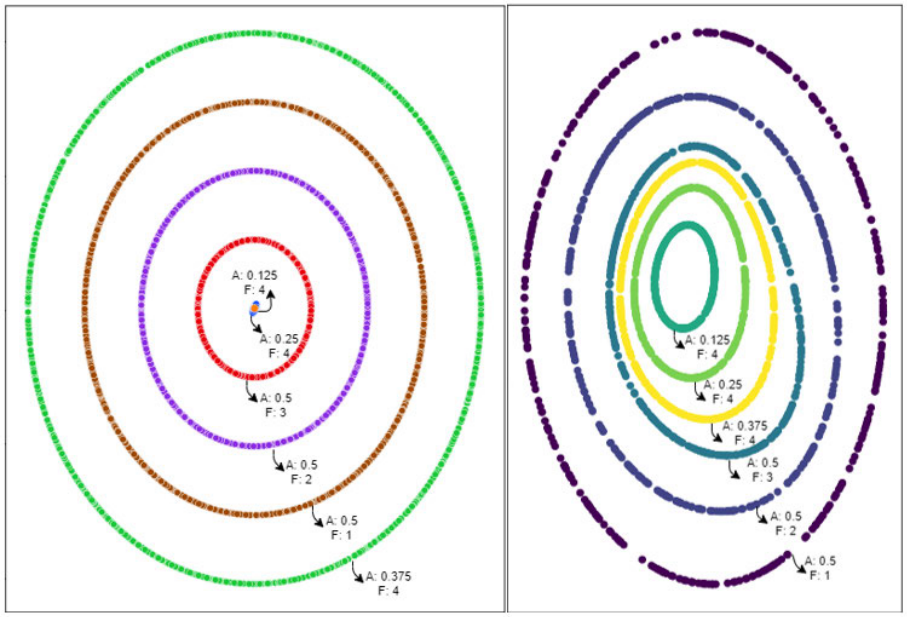

FIGURE 8. Simplified representations of sine wave sequences by PCA (left) and our model (right). It is shown that in our solution the organisation of the clusters is consistent according to the data while in PCA the organisation may not fit the nature of the data.

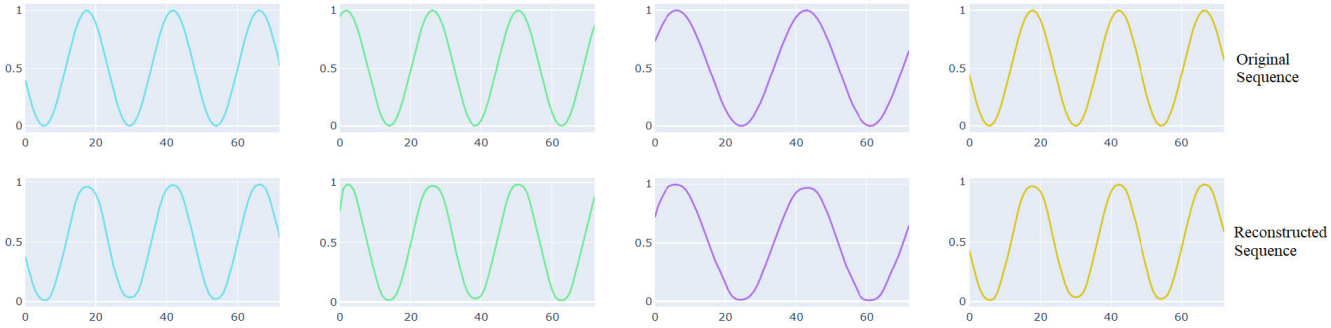

FIGURE 9. Comparison of reconstructed and original senoid samples.

(Wang *et al.* 2017b) with a main difference where the GAP layer is replaced with an attention layer.

- TWIESN: Time Warping Invariant Echo State Network, a variant of the Echo State Networks (ESN) proposed by [41] in which each timestep is projected in a space whose dimensions are inferred from the size of a reservoir. Then for each element, a Ridge classifier is trained to predict the class of each element in the time series.

It should be noted that in the previous section the importance of RNN for time series processing was highlighted whereas in this study TWIESN is presented as the only RNN to be compared. RNNs are generally applied for time series forecasting, however, when it comes to classification there are some drawbacks that emerge:

- This type of architecture is primarily designed to predict an output for each element in the time series [42].
- RNNs often suffer from the Vanishing Gradient problem due to long time series training [43].
- RNNs are considered difficult to train and parallelize, which leads to the avoidance of their use for computational reasons [44].

In our case the main objective is not classification but the treatment of the evolution in the series that represent the arrhythmias, which is why the application of other architectures was not considered.

Table 1 shows the performance of the different models for each class in terms of accuracy. Each entry in the table is the number of times an arrhythmia in a class was recognized by each model for the appropriate class. In addition, to illustrate the performance of each method, the ranking calculated by the Friedman method (ranking by range) for each dataset and the resulting averaged ranking are included.

TABLE 1. Accuracy of the different classifiers, 6 types of AF.

	Accuracy				
	Resnet	FCN	Encoder	TWIESN	RVAE
998na10	0.9681(2)	0.9867 (1)	0.9361(5)	0.9517(4)	0.9543(3)
998na30	0.9664(1)	0.8788(5)	0.9456(3)	0.9438(4)	0.9553(2)
998na180	0.9846(1)	0.9719(4)	0.9729(3)	0.9611(5)	0.9779(2)
999na10	0.9849(1)	0.9849(2)	0.9364(5)	0.9505(4)	0.9770(3)
999na30	0.9879(1)	0.9778(3)	0.9826(2)	0.9791(4)	0.9733(5)
999na180	0.9904(1)	0.9886(4)	0.9895(3)	0.9786(5)	0.9895(2)
	Summary Results				
Accuracy	0.9803	0.9647	0.9603	0.9607	0.9712
Average rank	1.166	3.166	3.500	4.333	2.833

It can be seen that the best classifier is Resnet, followed by our solution, labelled as RVAE. To extend the comparison between the different methods, post-hoc tests were carried out to detect significant differences in pairs between all the classifiers as recommended in [45]. Table 2 shows the family of hypotheses formulated to compare the classifiers ordered by the corresponding p-values. If the significance test yields a p-value lower than a predefined threshold (usually 0.05), then the difference is considered significant, therefore one model is declared superior to another. In this case only Resnet is significantly higher than the other models, which are FCN, Encoder and TWIESN if a significance level of 0.05 is considered since the p-values are below this threshold. The only solution to which it does not significantly exceed is ours. If the Bonferroni correction is considered, in which the number of comparisons is taken into account, the threshold which would have to be set is 0.05 divided by the number of comparisons, i.e. $0.05/6 = 0.0083$. Taking this value, Resnet would only be significantly higher than TWIESN. This is important to note because only TWIESN and our solution use RNNs, so it can be stated that our solution outperforms the best state of the art RNN classifier.

As a conclusion of this comparative study, it can be stated that our framework is capable of competing with the best time series classifiers reported up to 2019. Besides, the misclassification errors of our model correspond to arrhythmias that by their characteristics are located between classes similar to the one that really belongs, see Figure 7 right side: the classifier learns from that representation, so it can be assumed that failures are most likely due to the overlap of instances of a similar nature, which can also be interpreted as an estimate of the class of arrhythmia that most resembles its parameters or even as the possible future evolution that they will have, as we will address next.

C. VISUAL DIAGNOSIS

As discussed in previous sections, the latent properties of the VAE were prioritized to obtain a latent space whose characteristics were suitable for a simplified representation of the data. Figure 7 (right side) is the result of the latent representations obtained by the encoder for the training data and it can be understood as a projection of the 5 parameters that govern the arrhythmias simulation model. There are six clusters, which correspond to the six classes with which

TABLE 2. Family of hypotheses ordered by p-value.

i	hypothesis	$z = (R_0 - R_i)/SE$	p
1	Resnet vs TWIESN	3.465	0.0005
2	Resnet vs Encoder	2.556	0.0106
3	Resnet vs FCN	2.191	0.0285
4	Resnet vs RVAE	1.826	0.0679
5	RVAE vs TWIESN	1.640	0.1010
6	FCN vs TWIESN	1.275	0.2023
7	Encoder vs TWIESN	0.912	0.3618
8	RVAE vs Encoder	0.731	0.4648
9	FCN vs Encoder	0.366	0.7144
10	FCN vs RVAE	0.365	0.7151

the VAE was trained, labelled according to the two most relevant parameters of the simulated arrhythmias: α and β. The representations are organised according to the criticality of these two parameters.

The clinical interest lies in being able to project a real patient's data onto the latent space to find out which parameters of the model fit best. The procedure is quite intuitive; the arrhythmias are fed to the encoder, which predicts an output that will be the mean and variance of each one adapted to the distribution learned during the training. These two parameters are the axes that govern the latent space therefore, their codification in this space corresponds to a point with coordinates X = mean and Y = variance. In short, each point represents an arrhythmia from the training set. By projecting these dimensions on the learned map, their location on the clusters of arrhythmias that are present will give insight about the parameters that best define the patient's condition.

Figure 10 shows a projection (red dots) of two randomly selected patients from their pacemaker records. On the left side, it can be seen that the patient's projection falls into the group belonging to arrhythmias that have parameters $\alpha = 0.999$ and $\beta = 180$. Remember, α measures the speed of progression of arrhythmias and values of α close to 0.999 indicate a slow progression of AF. β indicates the average time between arrhythmias, in this case, it is more likely that those of this patient will occur at least every 180 days, so it is estimated that this is a patient that progresses positively without involving much risk.

As the map is organised, it is evident that the values of β are located from left to right from highest to lowest (180, 30, 10), which is equal to an organisation from lowest to highest

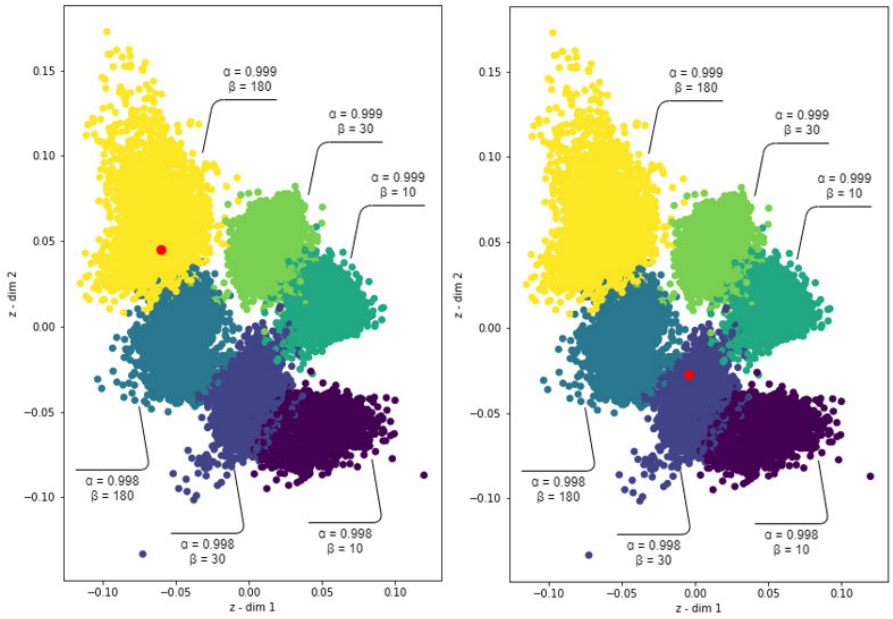

FIGURE 10. Projection of arrhythmias of actual patients.

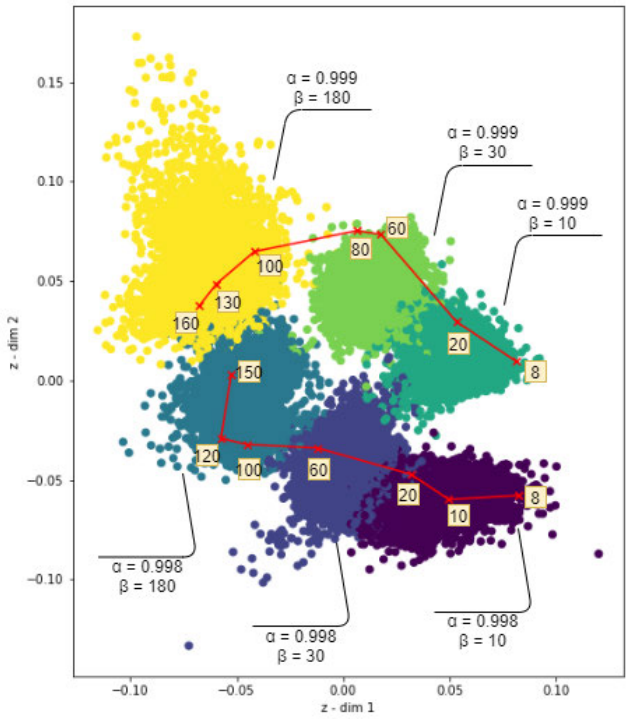

FIGURE 11. Projection of unseen simulated cases with known parameters. It can be seen that the arrhythmias are projected in the correct cluster. Besides, their location on the map is consistent according to their proximity to other clusters.

criticality as low values of β indicate short times between different episodes. On the other hand, the values of α are organised from top to bottom (999, 998), from less to more critical. This information can be used to facilitate a better interpretation of the map. The upper right zone denotes the less critical arrhythmias, while the lower right zone shows those arrhythmias that represent a very advanced stage of the disease. At the same time, the rest of the parameters of the simulation model during the generation of the training set have been varied randomly, which slightly influences the condition of the arrhythmias, therefore this property can give rise to the interpretation of arrhythmias between two clusters as an interpolation between the parameters of two classes.

This fact can be seen in the patient depicted on the right side of the same figure. This second case is located in the cluster with parameters $\alpha = 0.998$ and $\beta = 30$. Firstly, the parameter β is closer to 30, but due to its proximity to the lower-left group ($\beta = 180$), it can be understood that its evolution is on the way to reach 30, possibly a value between 180 and 30. Secondly, the most critical parameter, α, corresponds to a value of 0.998, which means that the evolution is closer to a permanent arrhythmia. This is not the most critical case, but it may need medical intervention in order to prevent future complications.

The organisation of the latent space reveals that the model is capable of setting apart the different values of α and β, allowing us to know if the condition of a certain patient evolves dangerously towards permanent AF. It is important to highlight the latent organisation obtained and its interpretability. As mentioned previously, the most dangerous arrhythmias are located on the lower right and those that do not suggest too much danger on the upper left. This evolution from one corner to the other can be interpreted as an interpolation of the parameters used to offer a diagnosis onto the latent space: α on the Y-axis and β on the X-axis. Figure 11 reinforces this idea: new simulated arrhythmias are projected onto the latent space by varying the parameters of the simulation model,

but unlike the process followed to generate the training data where the parameters λ_{GA}, $\lambda_A(t)$ and p_{AG} were randomly altered within a certain range, on this occasion they were left fixed. As a result, instances are represented as crosses and labelled according to the parameter β with which they were generated. The parameter α is omitted since the membership towards 0.998 or to 0.999 is evident. The projection on the map shows that the interpretation of the parameters of a given arrhythmia can be established according to the proximity to a specific cluster. That is, despite the fact that the first group characterizes those arrhythmias with parameters $\alpha = 0.999$ and $\beta = 10$, if an arrhythmia is located in the limit between this group and the one on its left it is very likely that it has an intermediate parameter β between both, (e.g. 20), or if an arrhythmia is located between a superior and an inferior group it would mean that the parameter α evolves dangerously towards values of 0.998. In this way, it is possible to know how the progression of a given arrhythmia could evolve.

VI. CONCLUDING REMARKS AND FUTURE WORK

We have described, trained and evaluated a recurrent VAE architecture based on Bidirectional LSTMs to assess the progression of time series by means of a graphic projection. We introduced a classifier to regularise the formation of the latent space and thus obtain a representation according to the nature of the data. The diagnosis of AF disease has been addressed with this model using intracardiac pacemaker records from actual patients and not only was an explainable diagnosis achieved but also our method was shown to be able to compete with solutions dedicated exclusively to time series classification, outperforming three of the four methods presented in terms of accuracy.

Lastly, the flexibility of the resulting model provides an opportunity to explore other future work contributions. Latent properties can be addressed in even more detail with recent architectures [46], [47] and the decoder, which has been discarded for this work, can be used for other tasks such as the detection of anomalies in data reconstruction or the prediction of the next time steps in the analysed time series.

REFERENCES

[1] A. Ng, "Sparse autoencoder," *CS294A Lect. Notes*, vol. 72, pp. 1–19, Jan. 2011.
[2] S. Rifai, G. Mesnil, P. Vincent, X. Muller, Y. Bengio, Y. Dauphin, and X. Glorot, "Higher order contractive auto-encoder," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Montreal, QC, Canada: Springer, 2011, pp. 645–660.
[3] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech enhancement based on deep denoising autoencoder," in *Proc. Interspeech*, 2013, pp. 436–440.
[4] D. P Kingma and M. Welling, "Auto-encoding variational bayes," 2013, *arXiv:1312.6114*. [Online]. Available: http://arxiv.org/abs/1312.6114
[5] N. Dilokthanakul, P. A. M. Mediano, M. Garnelo, M. C. H. Lee, H. Salimbeni, K. Arulkumaran, and M. Shanahan, "Deep unsupervised clustering with Gaussian mixture variational autoencoders," 2016, *arXiv:1611.02648*. [Online]. Available: http://arxiv.org/abs/1611.02648
[6] Z. Jiang, Y. Zheng, H. Tan, B. Tang, and H. Zhou, "Variational deep embedding: An unsupervised and generative approach to clustering," 2016, *arXiv:1611.05148*. [Online]. Available: http://arxiv.org/abs/1611.05148
[7] L. Yang, N.-M. Cheung, J. Li, and J. Fang, "Deep clustering by Gaussian mixture variational autoencoders with graph embedding," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6440–6449.
[8] K.-L. Lim, X. Jiang, and C. Yi, "Deep clustering with variational autoencoder," *IEEE Signal Process. Lett.*, vol. 27, pp. 231–235, 2020.
[9] X. Hou, L. Shen, K. Sun, and G. Qiu, "Deep feature consistent variational autoencoder," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 1133–1141.
[10] J. Walker, C. Doersch, A. Gupta, and M. Hebert, "An uncertain future: Forecasting from static images using variational autoencoders," in *Proc. Eur. Conf. Comput. Vis.* Philadelphia, PA, USA: Springer, 2016, pp. 835–851.
[11] A. Dosovitskiy and T. Brox, "Generating images with perceptual similarity metrics based on deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 658–666.
[12] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," 2015, *arXiv:1511.06349*. [Online]. Available: http://arxiv.org/abs/1511.06349
[13] I. V. Serban, A. Sordoni, R. Lowe, L. Charlin, J. Pineau, A. Courville, and Y. Bengio, "A hierarchical latent variable encoder-decoder model for generating dialogues," in *Proc. 21st AAAI Conf. Artif. Intell.*, 2017, pp. 3295–3301.
[14] Z. Yang, Z. Hu, R. Salakhutdinov, and T. Berg-Kirkpatrick, "Improved variational autoencoders for text modeling using dilated convolutions," 2017, *arXiv:1702.08139*. [Online]. Available: http://arxiv.org/abs/1702.08139
[15] O. Fabius and J. R. van Amersfoort, "Variational recurrent auto-encoders," 2014, *arXiv:1412.6581*. [Online]. Available: http://arxiv.org/abs/1412.6581
[16] D. Park, Y. Hoshi, and C. C. Kemp, "A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1544–1551, Jul. 2018.
[17] S. Suh, D. H. Chae, H.-G. Kang, and S. Choi, "Echo-state conditional variational autoencoder for anomaly detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 1015–1022.
[18] G. J. Padfield, C. Steinberg, J. Swampillai, H. Qian, S. J. Connolly, P. Dorian, M. S. Green, K. H. Humphries, G. J. Klein, R. Sheldon, M. Talajic, and C. R. Kerr, "Progression of paroxysmal to persistent atrial fibrillation: 10-year follow-up in the canadian registry of atrial fibrillation," *Heart Rhythm*, vol. 14, no. 6, pp. 801–807, Jun. 2017.
[19] P. K. D. Pramanik, B. K. Upadhyaya, S. Pal, and T. Pal, "Internet of Things, smart sensors, and pervasive systems: Enabling connected and pervasive healthcare," in *Healthcare Data Analytics and Management* (Advances in Ubiquitous Sensing Applications for Healthcare), N. Dey, A. S. Ashour, C. Bhatt, and S. J. Fong, Eds. New York, NY, USA: Academic, 2019, pp. 1–58.
[20] A. Bansal and R. Joshi, "Portable out-of-hospital electrocardiography: A review of current technologies," *J. Arrhythmia*, vol. 34, no. 2, pp. 129–138, 2018.
[21] M. P. Turakhia, M. Desai, H. Hedlin, A. Rajmane, N. Talati, T. Ferris, S. Desai, D. Nag, M. Patel, P. Kowey, J. S. Rumsfeld, A. M. Russo, M. T. Hills, C. B. Granger, K. W. Mahaffey, and M. V. Perez, "Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The apple heart study," *Amer. Heart J.*, vol. 207, pp. 66–75, Jan. 2019.
[22] D. Calvo, J. Rubín, D. Pérez, and J. Jalife, "Spectral analysis of electrograms in a substrate modified by radiofrequency ablation reveals similarities between organized and disorganized atrial rhythms," *Heart Rhythm*, vol. 11, no. 12, pp. 2306–2309, Dec. 2014.
[23] G. N. Kay, K. A. Ellenbogen, M. Giudici, M. M. Redfield, L. S. Jenkins, M. Mianulli, and B. Wilkoff, "The ablate and pace trial: A prospective study of catheter ablation of the av conduction system and permanent pacemaker implantation for treatment of atrial fibrillation," *J. Interventional Cardiac Electrophysiol.*, vol. 2, no. 2, pp. 121–135, 1998.
[24] C. D. Swerdlow, G. Kalahasty, and K. A. Ellenbogen, "Implantable cardiac defibrillator lead failure and management," *J. Amer. College Cardiol.*, vol. 67, no. 11, pp. 1358–1368, Mar. 2016.
[25] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Briefings Bioinf.*, vol. 19, no. 6, pp. 1236–1246, Nov. 2018.
[26] M. R. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI," *Med. Image Anal.*, vol. 30, pp. 108–119, May 2016.

[27] J. Eom, S. Kim, and B. Zhang, "AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2465–2479, May 2008.
[28] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ecg features for screening paroxysmal atrial fibrillation patients," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 12, pp. 2095–2104, 2018.
[29] S. P. Shashikumar, A. J. Shah, Q. Li, G. D. Clifford, and S. Nemati, "A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Feb. 2017, pp. 141–144.
[30] N. Costa, J. Fernández, I. Couso, and L. Sánchez, "Graphical analysis of the progression of atrial arrhythmia using recurrent neural networks," *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, pp. 1567–1577, 2020.
[31] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[32] J. Fernández, J. Velasco, and L. Sánchez, "Detection of cardiac arrhythmias through singular spectrum analysis of a time-distorted EGM signal," in *Proc. Int. Joint Conf. SOCO CISIS ICEUTE León, Spain, September 6–8, 2017*, H. Pérez García, J. Alfonso-Cendón, L. Sánchez González, H. Quintián, and E. Corchado, Eds. Cham, Switzerland: Springer, 2018, pp. 137–146.
[33] D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling, "Semi-supervised learning with deep generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 3581–3589, 2014.
[34] F. Berkhahn, R. Keys, W. Ouertani, N. Shetty, and D. Geißler, "Augmenting variational autoencoders with sparse labels: A unified framework for unsupervised, semi-(un)supervised, and supervised learning," 2019, *arXiv:1908.03015*. [Online]. Available: http://arxiv.org/abs/1908.03015
[35] J. Bergstra, D. Yamins, and D. Cox, "Hyperopt: A Python library for optimizing the hyperparameters of machine learning algorithms," in *Proc. 12th Python Sci. Conf.*, 2013, p. 20.
[36] L. N. Smith, "Cyclical learning rates for training neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 464–472.
[37] F. Chollet, *Keras*. Mountain View, CA, USA: François Chollet, 2015.
[38] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: A review," *Data Mining Knowl. Discovery*, vol. 33, no. 4, pp. 917–963, Jul. 2019.
[39] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 1578–1585.
[40] J. Serrà, S. Pascual, and A. Karatzoglou, "Towards a universal neural network encoder for time series," in *Proc. CCIA*, Oct. 2018, pp. 120–129.
[41] P. Tanisaro and G. Heidemann, "Time series classification using time warping invariant echo state networks," in *Proc. 15th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2016, pp. 831–836.
[42] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognit. Lett.*, vol. 42, pp. 11–24, Jun. 2014.
[43] R. Pascanu, T. Mikolov, and Y. Bengio, "Understanding the exploding gradient problem," 2012, *arXiv:1211.5063*. [Online]. Available: https://arxiv.org/abs/1211.5063
[44] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1310–1318.
[45] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Jan. 2006.
[46] A. V. D. Oord, O. Vinyals, and K. Kavukcuoglu, "Neural discrete representation learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6306–6315.
[47] K. Gregor, G. Papamakarios, F. Besse, L. Buesing, and T. Weber, "Temporal difference variational auto-encoder," 2018, *arXiv:1806.03107*. [Online]. Available: http://arxiv.org/abs/1806.03107

NAHUEL COSTA received the B.Sc. degree in computer engineering in information technology and the M.Sc. degree in computer science from the University of Oviedo, Spain, in 2019 and 2021, respectively, where he is currently pursuing the Ph.D. degree in artificial intelligence with the Computer Science Department. He is also working as a Researcher with the Computer Science Department, University of Oviedo.

LUCIANO SÁNCHEZ (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in electronic engineering from the University of Oviedo, Spain, in 1991 and 1994, respectively. He is currently a Full Professor with the Department of Computer Science, University of Oviedo. He is the author of more than 80 international journal articles and more than 100 conference papers and book chapters. His research interests include the theoretical study of algorithms for mathematical modeling and intelligent data analysis, and the application of these techniques to practical problems of industrial modeling, signal processing and condition monitoring, with special interest in the study of low quality data and fuzzy information. He received the IEEE Outstanding Paper Award in 2013 IEEE International Conference on Fuzzy Systems, Hyderabad, India, and the 2013 Rolls-Royce Deutschland Engineering Innovationspreis, Berlin, Germany.

INÉS COUSO (Member, IEEE) received the M.Sc. and Ph.D. degrees in mathematics from the University of Oviedo, Spain, in 1995 and 1999, respectively. She has been a Visiting Professor at several institutes and laboratories, such as the Institut de Recherche en Informatique de Toulouse, Université Paul Sabatier, Toulouse, France, and the Laboratoire de Informatique, de Robotique et de Microélectronique de Montpellier, Université Montpellier II. She was a Professeur invitée 1er classe with the CNRS, France, in 2008. She was a Professeur invitée with the Université Montpellier 2, France, in 2011. She is currently a Full Professor in statistics and operations research with the University of Oviedo. She was named Excellence Advanced Researcher by the Centre International de Mathematique et d'Informatique de Toulouse (CIMI), in 2015. She has written 80 articles in peer-reviewed journals and more than 100 book chapters and communications to national and international conferences. She is also a Senior Editor of the *International Journal of Approximate Reasoning* and an Area Editor of the journal *Fuzzy Sets and Systems*.