Machine Learning approaches used to empower predictivities of hiPSC-CM assays

Haibo Liu [12] Sylvain Bernasconi and Christophe Bleunven [1] Damiano Lombardi and Muriel Boulakia [2]

[1]NOTOCORD SYSTEMS [2]Inria

Neural Networks in high-throughput screening

Considering that MEA datasets from hiPSC-CM assays are usually very large, most researchers and existing tools mainly focus on one electrode per well (the "golden electrode") to extract these biomarkers, whereas valuable information may reside in the other and not considered electrodes. In this work, we propose to test Artificial Neural Network (ANN) methods to be able to benefit from using all the data to have more information in order to answer some more complex questions. More specifically, given the extracellular field potential (FP) recorded by MEA, we aimed to determine whether a given chemical compound alters the electrical activity of cardiomyocytes by disrupting the normal behaviour of the hERG channels.

Artificial Neural Network used to detect hERG channel blockers

This problem can be translated as a binary classification problem. It means, given new data, the trained ANN categorizes it into one of two classes (No-hERG blocker or hERG blocker). An ANN consists of artificial neurons in a highly interconnected network structure that can send information to each other to process data like the way the human brain works as shown in Figure 1.

To do so, this study considered four different ANN methods and compared them in terms of their accuracy and computational costs.

- **Multilayer perceptron (MLP)**: input biomarkers
- **Univariate 1-dimensional Convolutional Neural Network (1D-CNN)**: input a single beat of signal
- **Multivariate 1D-CNN**: input a matrix which is a combination of a few beats
- **2-D CNN**: input images (we can transform signals into images)

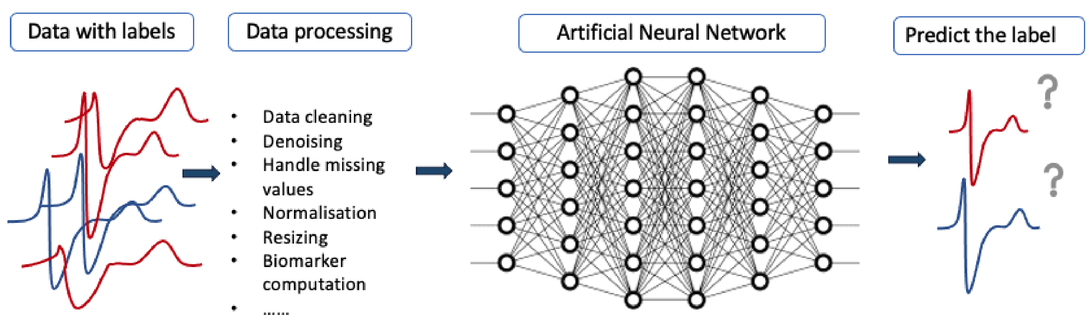

Figure 1. Explanation of ANN

Training data set

- **Training data from:** 7 hERG blockers (loratadine, ibutilide, Mexiletine, Droperidol, Chlorpromazine, Clozapine, and Dofetilide) and No-hERG blockers (Diltazem plus control vehicles)
- **Testing data from:** 4 hERG blockers (Clarithromycine, Cisapride, Bepridil, Azimilide)

Results of hERG channel blocker prediction

The results are summarised in Table 1, we can see that MLP and Multivariate 1D-CNN provided the most promising prediction.

- **MLP** is a good choice when we do not mind the cost of computing biomarkers
- **Multivariate 1D-CNN** is a good option when we would like to save costs from the computation of bio-markers

Table 1. Classification results of the tested ANNs

Metrics	Classifier			
	MLP	Univariate 1D-CNN	Multivariate 1D-CNN	2D-CNN
Accuracy	**98.33%**	64.63%	**99.13%**	84.42%
Data Processing Cost	**Very High**	Very Low	**Very Low**	Low
Training Cost	**20 minutes**	5.8 hours	**1.2 hours**	6.7 hours

For more details, please check the published article in [1].

Want to know more? See us on the Instem booths or leave your card

Machine Learning in disease study

On the second topic, based on collaboration with the University of Antwerp, we employed Machine Learning (ML) methods to study Brugada syndrome using patch-clamp recordings (Action Potential, AP) obtained from patient-derived iPSC-CM.

- **Difficult**: some patients expressed a more severe syndrome than others, while they all carried the same mutation.
- **Our goal**: investigate if ML approaches could identify patients with mild and severe syndromes.

Analytical challenges arising from variability in AP recordings

Because there are big variations between different clones and cell lines in both control and patient AP.

- **Both Atrial like and Ventricular like APs**
- **Not standard AP shapes**

Analysing them all mixed becomes a big challenge. The method we used is called density-based spatial clustering of applications with noise (DBSCAN). It allows us to group APs by their similarities. In this way, we can separate the whole records by their characteristics and then analyse the control and patients' AP by groups.

Clustering method to handle the analytical challenges

The APs from control cells were divided into 36 clusters, here we put a few examples in Figure 2.

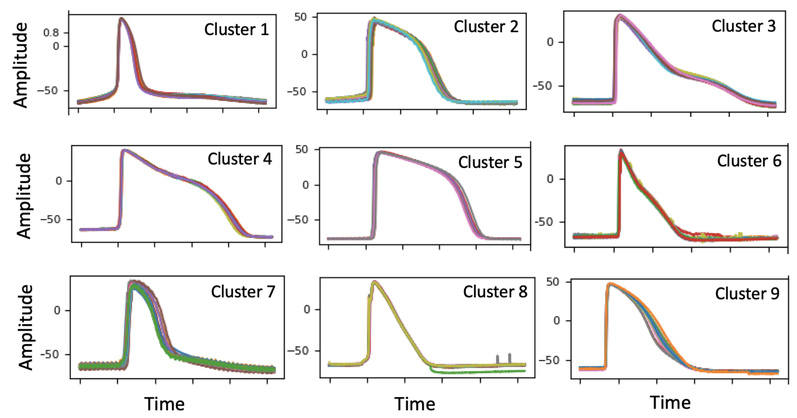

Figure 2. Clustering AP of control cell lines

Identify patients compare with control AP

1. Given a patient AP, compute the distance to the centre of each cluster in the control data.
2. Assign control cluster numbers to patients' AP by selecting the minimum distance.

A patient's AP was considered an outlier if there was no control cluster similar to it. In principle, more outliers exist, suggesting more severe syndromes may exist because more APs do not look similar to any control AP. The results summarised in Table 2 show that

- **Patient 1 has more similar AP compared to the control AP.**
- **Patient 1 has fewer outliers compared to the control AP.**

We can make a suggestion that Patient 1 may have a less severe syndrome compared to Patient 2 which agrees with the prior information from our collaborator. In conclusion, the clustering method can be used to group data when we have difficulties dealing with data with variations.

Table 2. Patients AP compare with control AP

	Very Similar AP	Outliers
Patient 1	52.61%	12.82%
Patient 2	34.29%	25.82%

References

[1] Haibo Liu, Tessa De Korte, Sylvain Bernasconi, Christophe Bleunven, Damiano Lombardi, and Boulakia Muriel. Artificial neural network comparison on herg channel blockade detection. *International Journal of Computer Applications*, 184(14):1–9, May 2022.

Funded by the EU H2020 Programme MSCA (GA 858070)