Title	Enhance Categorisation Of Multilevel High-Sensitivity Cardiovascular Biomarkers From Lateral Flow Immunoassay Images Via Neural Networks And Dynamic Time Warping
Authors(s)	Jing, Min, MacNamee, Brian, McLaughlin, Donal, et al.
Publication date	2020-10-28
Publication information	Jing, Min, Brian MacNamee, Donal McLaughlin, and et al. "Enhance Categorisation Of Multilevel High-Sensitivity Cardiovascular Biomarkers From Lateral Flow Immunoassay Images Via Neural Networks And Dynamic Time Warping." IEEE, 2020.
Conference details	The 27th IEEE International Conference on Image Processing, Abu Dhabi, United Arab Emirates (held online due to coronavirus outbreak), 25-28 October 2020
Publisher	IEEE
Item record/more information	http://hdl.handle.net/10197/24610
Publisher's statement	© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Publisher's version (DOI)	10.1109/ICIP40778.2020.9190827

ENHANCE CATEGORISATION OF MULTILEVEL HIGH-SENSITIVITY CARDIOVASCULAR BIOMARKERS FROM LATERAL FLOW IMMUNOASSAY IMAGES VIA NEURAL NETWORKS AND DYNAMIC TIME WARPING

Min Jing[1], Brian Mac Namee[2], Donal McLaughlin[3], David Steele[4], Sara McNamee[1],Patrick Cullen[1]
Dewar Finlay[1], James McLaughlin[1]

[1] Nanotechnology and Integrated BioEngineering Centre (NIBEC), Ulster University, United Kingdom
[2] School of Computer Science, University College Dublin, Republic of Ireland
[3] Department of Physics, University College London, United Kingdom
[4] Biocolor Ltd, United Kingdom

ABSTRACT

Lateral Flow Immunoassays (LFA) are low cost, rapid and highly efficacious Point-of-Care devices. Traditional LFA testing faces challenges to detect high-sensitivity biomarkers due to low sensitivity. Unlike most approaches based on averaging image intensity from a region-of-interest (ROI), this paper presents a novel system that considers each row of an LFA image as a time series signal and, consequently, does not require the detection of ROI. Long Short-Term Memory (LSTM) networks are used to classify LFA data obtained from multilevel high-sensitivity cardiovascular biomarkers. Dynamic Time Warping (DTW) was incorporated with LSTM to align the LFA data from different concentration levels to a common reference before feeding the distance maps into an LSTM network. The LSTM network outperforms other classifiers with or without DTW. Furthermore, performance of all classifiers is improved after incorporating DTW. The positive outcomes suggest the potential of the proposed methods for early risk assessment of cardiovascular diseases.

Index Terms— Lateral Flow Immunoassays, Long Short-Term Memory, Dynamic Time Warping, high-sensitivity cardiovascular biomarkers, classification

1. INTRODUCTION

There is a growing demand for a range of portable, rapid and low-cost biosensing devices for the early detection of cardiovascular disease (CVD). Lateral Flow Immunoassays (LFA) have attracted increased attention recently, but the low sensitivity of LFA limits their ability to detect cardiovascular biomarkers, such as high-sensitivity C-Reactive Protein (hs-CRP) tests performed over a lower range (from 0.5mg/L to 10 mg/L), which can be used for early risk assessment of CVD [1]. The enzyme-linked immunosorbent assay (ELISA) is more sensitive but more time consuming than LFA [2]. Improvements can be made by developing high sensitivity assays [3], improving labeling strategies, enhancing the optical and electrochemical transducers and exploring the evolution of recognition [4]. However, most of these approaches require either external equipment, high-cost reagents, or complicated fabrication with multistep procedures.

Smartphone-based LFA testing approaches have been reported recently [5] [6] but most are focused on binary classification via Support Vector Machine (SVM) based on mean of image intensity from a recognised region-of-interest (ROI). The performance of these approaches can be affected by image quality and the accuracy of ROI detection. Very few studies [7] [8] have applied neural networks to LFA scenarios. Deep Belief Networks (DBN) were applied in [7] for LFA testing based on human chorionic gonadotropin, in which DBN was used to improve the efficiency of region of interest (ROI) detection rather than classification. A Multi-Layer Perceptron (MLP) neural network was used in [8] for drugs-of-abuse detection based on image intensity to assess saliva content rather than blood-based high-sensitivity biomarkers.

This paper presents a novel system considering LFA image as time series that captures the temporal information richer than image intensity, which aims to enhance the performance involving multilevel or multiplex LFA testing [9] via incorporating Dynamic Time Warping (DTW) [10] with Long Short-Term Memory (LSTM) networks [11]. The rest of this paper is organised as follows. The data details are explained in Section 2. In Section 3, the framework is described including the LSTM and arrangement of DTW outputs for LSTM. The experimental results are given in Section 4 with the conclusion in Section 5.

2. LFA IMAGE DATA

For LFA image data commonly obtained from the scanners or smartphone cameras, the performance of testing can be affected by image quality since data are acquired under ambi-

ent lighting. In this study, a CMOS reader system has been designed, in which LFA image data were obtained from an opaque box under controlled lighting conditions. Fig. 1 gives examples of a set of LFA strip images from eight hs-CRP concentration levels in a range from 0.05 mg/L to 5 mg/L, which is aligned with clinically actionable categories for early risk assessment of CVD [1]. The LFA images were captured at a fixed time point (also known as an endpoint assay), following 'completion' of the lateral flow assay. It can be seen from Fig.1 that intensity of the Test line (T-line) changes according to the concentration level of the target biomarkers. The Control line (C-line) captures any particle therefore always appears regardless of the presence of the target analyte. The flow direction indicates that the sample flows from T-line to C-line via capillary action. The images contain not only particular spatial phenomenon (i.e., the leading edge is stronger than the trailing edge), but also valuable time-dependent information that arises from the interplay between biomarker and labelled conjugate antibodies as it develops over time. It is also noticed that the position of the T-line varies in each image. For approaches based on image intensity, the performance relies on the accuracy of detection of ROI (the T-line area). This study considers the LFA data along the sample flow direction as time series signals which not only captures the temporal information but also no ROI detection is needed.
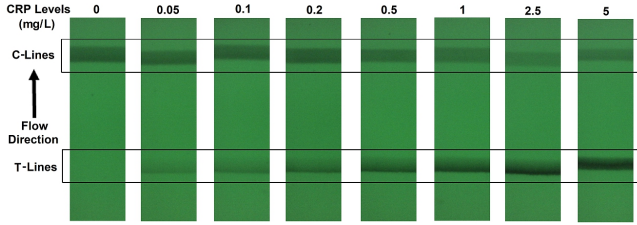

Fig. 1. Examples of LFA images from eight hs-CRP levels.

3. METHODS

3.1. System Overview

A block diagram of the proposed framework is given in Fig. 2. As mentioned above, the C-line does not change according to concentration levels, the half of an image containing the T-line is directly selected for testing. (In a fully integrated analysis system the formation of the C-line would be used as a quality control check to ensure the assay has performed correctly). An LSTM network can be directly applied to LFA signals but better performance is expected with additional features. Since LFA time series reveal the change of intensity with time steps (examples are given in Fig. 4), DTW was applied to align the time series from all levels to one reference LFA image from level-8. The distance maps from DTW capture the feature of alignment, which are arranged as input sequences for LSTM, followed by a fully connected layer, a softmax layer and an output layer for sequence classification.
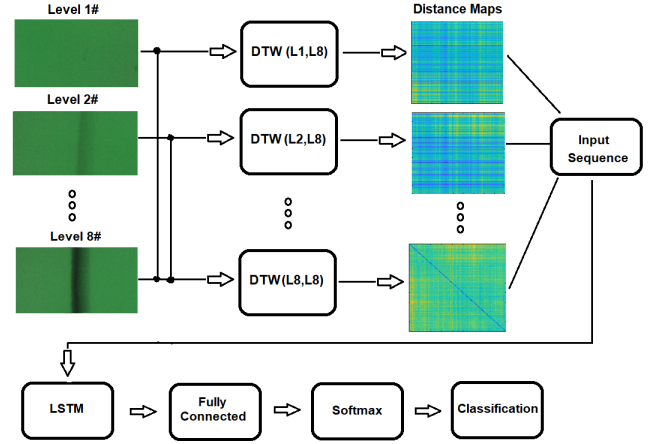

Fig. 2. The block diagram for the proposed framework.

3.2. Incorporating DTW and LSTM

LSTM is one type of Recurrent Neural Networks (RNN) that can learn long-term dependencies in time series data. LSTM networks have been applied successfully to speech recognition [12], language modelling [13] and ECG arrhythmia detection [14]. LSTM networks update the information for the current state based on the previous state via different gates. A block diagram of the LSTM is given in Figure 3. Given a time series sequence $\mathbf{X}$ with k features of length N, the input sequence for LSTM at the current time step t can be presented as a vector $\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_k(t)]^T$, where T denotes the transpose operation. A cell state $\mathbf{c}_t$ and the hidden state (output state) $\mathbf{h}_t$ are updated based on each at the previous time step $(\mathbf{c}_{t-1}, \mathbf{h}_{t-1})$. There are four components to control the system, i_t, f_t, g_t and o_t, which denote the input gate, forget gate, cell candidate and output gate, respectively. More details on how the network is updated via each gate can be found in [11].
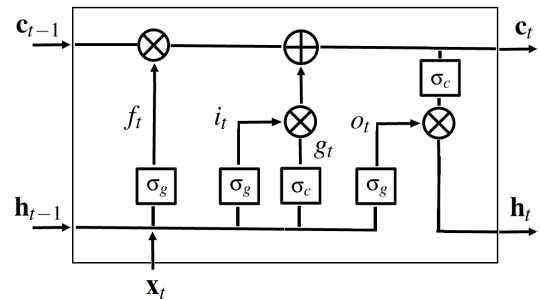

Fig. 3. Illustration of LSTM model.

In this study, DTW was applied to align the LFA time series to a common reference first then the distance maps obtained from DTW were fed into LSTM for classification. According to DTW [10], given two one-dimensional feature vectors, $\mathbf{a} = [a_1, a_2, ..., a_i]$ and $\mathbf{b} = [b_1, b_2, ..., b_j]$, which have i and j samples respectively. A time warping function is defined as

$$F = [p_1, p_2, ..., p_l] \tag{1}$$

where p is a sequence of points $p(l) = (i(l), j(l))$, which maps the time axis of feature **a** onto that of feature **b**. The difference between two features a_i and b_j can be measured by the Euclidean distance function $d(p) = d(i, j) = ||a_i - b_j||$. When there is no difference, the distance becomes zeros at the diagonal line ($i = j$). The DTW algorithm finds the optimal solutions by minimising the distances between corresponding points.

For half LFA image, the number of row is 450 pixels and number of column is 800 pixels, in which 450 is the width of LFA strip and 800 is the length across the strip (along the flow direction). Each row of images is considered as a time series (with 800 time-steps) since they contain the information that arises as a result of temporo-spatial interactions throughout the assay time (via the gradual accumulation of label conjugate particles). To apply DTW, each row of LFA images at different concentration levels was aligned to the corresponding row from a reference, therefore, the distance map is a matrix with dimension of 450×450. To ensure all data is compared with the same reference, a fixed LFA image from level-8 was selected as a reference. It is noticed that different number of input sequences can be obtained by dividing the distance map into a number of mini maps, which may affect the performance. A sequence for LSTM can be presented as:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & ... & x_{1N} \\ x_{21} & x_{22} & ... & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k1} & x_{k2} & ... & x_{kN} \end{pmatrix} \quad (2)$$

where $\mathbf{X}$ is the sequence based on a mini distance map with dimension k and $N = 450$. Then the number of sequences from one distance map is $450/k$. The dependence with the dimension of distance map and the number of sequences was investigated and the results are presented in the following sections.

4. EXPERIMENTAL RESULTS

The experiments were carried out based on LFA data obtained from eight hs-CRP concentration levels and each level has 30 LFA images (hence 240 images in total). Each image contains 450 time series so the total number of time series available is $450 \times 30 \times 8 = 108,000$. For all experiments, a holdout data partition was used, in which 90% were randomly selected for training and the remaining 10% for testing. (Different data partition and cross-validation will be considered in future work). The number of input sequences for training and testing under different arrangements are given in Table 1, in which the left column shows the number of dimensions and sequences that can be obtained from one distance map. The accuracy was defined as: sum (Predict = Test)/(Number of Test). The number of epochs, batch size and iteration rate was empirically set to 30, 32 and 0.01, respectively. The distance maps were rescaled to a range of $[0, 1]$ before being normalised by zscore and fed into the LSTM network. All numerical aspects of the experimentation were conducted using MATLAB2019b.

Table 1. Size of Sequences for Training and Testing

(Dimension, Sequences)	Training	Testing
(10, 45)	9720	1080
(15, 30)	6480	720
(30, 15)	3240	360
(45, 10)	2160	240
(90, 5)	1080	120

4.1. Results from DTW

An example of one set of original LFA time series from eight hs-CRP levels is shown in Fig.4 (a), which shows that the intensity changes according to the concentration levels. Fig. 4(b) shows the results after DTW, in which the original signals from each level were aligned to the one at level-8 (which remains itself as the reference). It is noticed that the length of the signals changed as DTW stretches the original data to match the reference. The examples of two distance maps from DTW by aligning LFA level-1 and level-8 data with the reference (level-8) are given in Fig. 5 (a) and (b) respectively, in which the distances are rescaled in the range of $[0, 1]$. It can be seen in Fig. 5 (b) that the values along the diagonal line are zeros because the rows from level-8 are aligned with themselves.

4.2. Dependence with Input Dimension

The results based on different settings for feature dimension in distance maps and the number of hidden layers in LSTM are provided in Fig. 6. The dimensions were set as 10, 15, 30, 45 and 90, and the number of hidden layers were 10, 25, 50, 100 and 150. The results show the best performance is achieved when the input dimension is 15 and number of hidden layers is 100, but 150 appears to give overall good performance for all cases. Same experiment was carried for LFA time series, the best setting for dimension is 45 and hidden layers 150.

4.3. Comparison of Classification

The classification by the proposed method was compared to five classifiers including SVM, K-Nearest Neighbours (KNN), Linear Discriminant Analysis (LDA), Decision Tree (DT) and Naive Bayes (NB). For fair comparison the same data partition was used to test all algorithms and same size of sequences as shown in Table 1. For LSTM, the number of hidden layers was selected as 150 (based on the evaluation in section 4.2). For other classifiers the multiclass models were trained based on the error-correcting output codes (ECOC) model [15]. The hyperparameters for each classifier were obtained via minimising five-fold cross-validation loss by varying parameters while running the automatic hyperparameter optimisation in Matlab. For SVM, the options for optimisation include "one vs one" or "one vs all", parameters for box constraint and kennel scale. For KNN, different distance functions and the number of neighbours were tested.
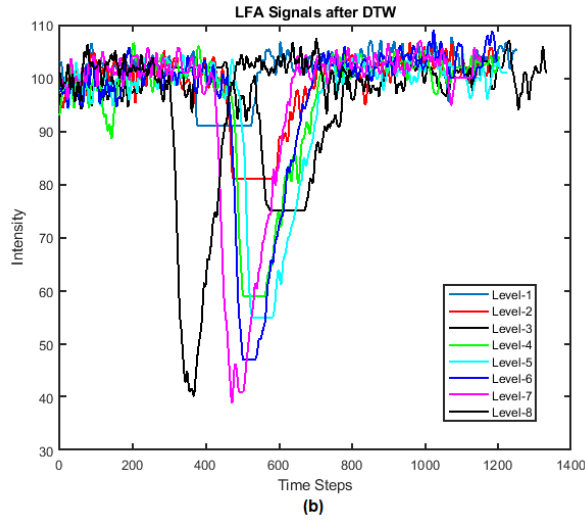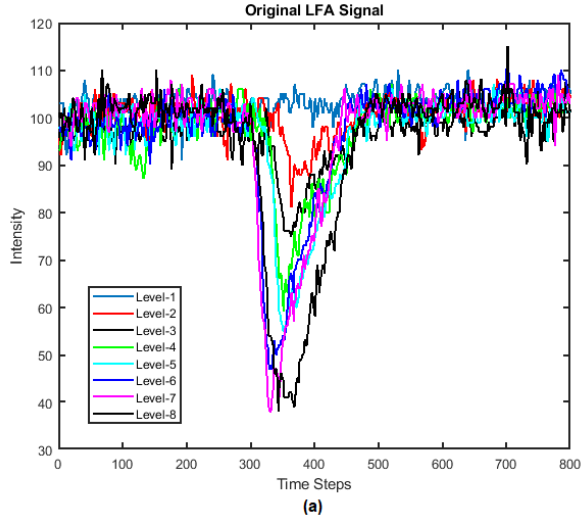
Fig. 4. LFA time series from eight concentration levels: (a) original signals; (b) signals after DTW.
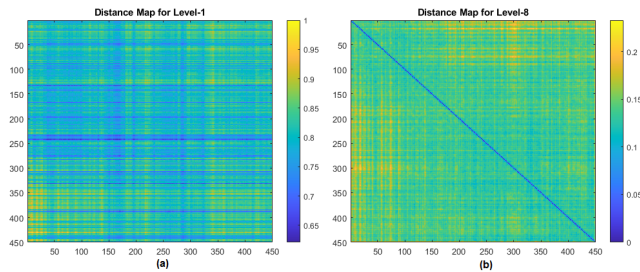

Fig. 5. Examples of two distance maps obtained from DTW by aligning LFA data with the reference (level-8): (a) level-1 and (b) level-8.

For LDA, the parameters for Delta (linear coefficient threshold) and Gamma (amount of regularisation) were considered. For DT, the parameter of leaf size was varied and the best was used for classification.

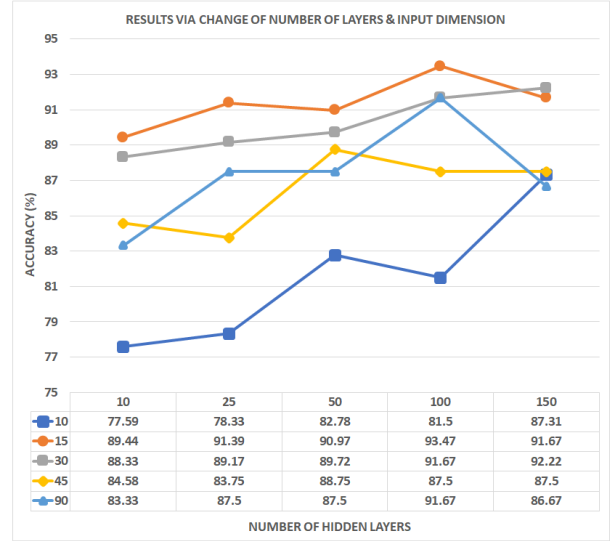The results based on LFA and LFA + DTW are given in

Fig. 6. The classification results based on different dimensions and number of hidden layers.

Table 2 and Table 3, respectively. It can be seen from Table 2, that for raw LFA time series data, LSTM clearly outperforms other classifiers. Table 3 shows that the performances for all classifiers are improved after incorporating DTW and the results from LSTM remain better than the rest.

Table 2. Classification based on LFA only

	Accuracy (%)					
Classifiers	(10,45)	(15,30)	(30,15)	(45,10)	(90,5)	Average
SVM	43.15	43.47	45.56	44.17	60.83	47.43
KNN	39.26	44.58	42.22	47.50	50.00	44.71
LDA	42.13	35.00	35.83	35.42	56.67	41.01
DT	51.57	45.97	42.22	52.92	57.50	50.03
NB	46.67	40.00	40.56	41.25	46.67	43.03
LSTM	**76.30**	**73.30**	**73.33**	**78.75**	**76.67**	**75.67**

Table 3. Classification based on LFA+DTW

	Accuracy (%)					
Classifiers	(10,45)	(15,30)	(30,15)	(45,10)	(90,5)	Average
SVM	84.54	81.53	84.17	81.25	79.17	82.13
KNN	70.46	74.17	80.56	76.67	79.17	76.21
LDA	81.39	84.03	83.33	80.83	79.17	81.75
DT	67.69	73.33	80.83	74.58	80.00	75.29
NB	60.65	61.81	66.94	67.08	70.00	65.30
LSTM	**85.02**	**82.92**	**85.00**	**86.25**	**85.93**	**85.01**

5. CONCLUSIONS

A novel system has been developed to enhance the detection of multilevel cardiovascular biomarker in LFA testing via incorporation of DTW and LSTM. Apart from no need for ROI detection, considering LFA image as the time series provides a new perspective for LFA data analysis and captures the valuable temporal information richer than image intensity. The outcomes based on the hs-CRP level below 5mg/L are encouraging, which suggest the potential of the proposed system not only for early risk assessment of CVD but also more complex LFA applications than binary classification in the future.

6. REFERENCES

[1] Paul M Ridker, Charles H Hennekens, Julie E Buring, and Nader Rifai, "C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women," *New England Journal of Medicine*, vol. 342, no. 12, pp. 836–843, 2000.

[2] Lucie Fojtíková, Anna Šuláková, Martina Blažková, Barbora Holubová, Martin Kuchař, Petra Mikšátková, Oldřich Lapčík, and Ladislav Fukal, "Lateral flow immunoassay and enzyme linked immunosorbent assay as effective immunomethods for the detection of synthetic cannabinoid jwh-200 based on the newly synthesized hapten," *Toxicology reports*, vol. 5, pp. 65–75, 2018.

[3] José L Torres and Paul M Ridker, "High sensitivity c-reactive protein in clinical practice," *American Heart Hospital Journal*, vol. 1, no. 3, pp. 207–211, 2003.

[4] Wing Cheung Mak, Valerio Beni, and Anthony PF Turner, "Lateral-flow technology: From visual to instrumental," *TrAC Trends in Analytical Chemistry*, vol. 79, pp. 297–305, 2016.

[5] Evgeni Eltzov, Sarah Guttel, Adarina Low Yuen Kei, Prima Dewi Sinawang, Rodica E Ionescu, and Robert S Marks, "Lateral flow immunoassays–from paper strip to smartphone technology," *Electroanalysis*, vol. 27, no. 9, pp. 2116–2130, 2015.

[6] Daniel Quesada-González and Arben Merkoçi, "Mobile phone-based biosensing: An emerging "diagnostic and communication" technology," *Biosensors and Bioelectronics*, vol. 92, pp. 549–562, 2017.

[7] Nianyin Zeng, Zidong Wang, Hong Zhang, Weibo Liu, and Fuad E Alsaadi, "Deep belief networks for quantitative analysis of a gold immunochromatographic strip," *Cognitive Computation*, vol. 8, no. 4, pp. 684–692, 2016.

[8] Adrian Carrio, Carlos Sampedro, Jose Sanchez-Lopez, Miguel Pimienta, and Pascual Campoy, "Automated low-cost smartphone-based lateral flow saliva test reader for drugs-of-abuse detection," *Sensors*, vol. 15, no. 11, pp. 29569–29593, 2015.

[9] Laura Anfossi, Fabio Di Nardo, Simone Cavalera, Cristina Giovannoli, and Claudio Baggiani, "Multiplex lateral flow immunoassay: an overview of strategies towards high-throughput point-of-need testing," *Biosensors*, vol. 9, no. 1, pp. 2, 2019.

[10] Hiroaki Sakoe and Seibi Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE transactions on acoustics, speech, and signal processing*, vol. 26, no. 1, pp. 43–49, 1978.

[11] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[12] Santiago Fernández, Alex Graves, and Jürgen Schmidhuber, "An application of recurrent neural networks to discriminative keyword spotting," in *International Conference on Artificial Neural Networks*. Springer, 2007, pp. 220–229.

[13] Rafal Jozefowicz, Oriol Vinyals, Mike Schuster, Noam Shazeer, and Yonghui Wu, "Exploring the limits of language modeling," *arXiv preprint arXiv:1602.02410*, 2016.

[14] Artzai Picon, Unai Irusta, Aitor Álvarez-Gila, Elisabete Aramendi, Felipe Alonso-Atienza, Carlos Figuera, Unai Ayala, Estibaliz Garrote, Lars Wik, Jo Kramer-Johansen, et al., "Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia," *PloS one*, vol. 14, no. 5, pp. e0216756, 2019.

[15] Thomas G Dietterich and Ghulum Bakiri, "Solving multiclass learning problems via error-correcting output codes," *Journal of artificial intelligence research*, vol. 2, pp. 263–286, 1994.