Patient-Specific Heartbeat Classification Based on I-Vector Adapted Deep Neural Networks

Sean Shensheng Xu, Man-Wai Mak, Chi-Chung Cheung
Dept. of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong SAR of China

***Abstract*—Automatic heartbeat classification from electrocardiogram (ECG) signals is important for diagnosing heart arrhythmias. A main challenge in ECG classification is the variability of ECG signals across patients. This paper proposes a patient-specific heartbeat classifier to address the inter-patient variations in ECG signals. Inspired by the success of identity vectors (i-vectors) in speech and speaker recognition, we extracted one i-vector from five minutes of ECG data for each patient and applied it to adapt a patient-independent deep neural network (DNN) to a patient-specific DNN, namely i-vector adapted patient-specific DNN (iAP-DNN). Evaluations on the MIT-BIH arrhythmia database show that the iAP-DNN is able to classify raw ECG signals of the corresponding patient into normal heartbeats and different types of arrhythmias and that it outperforms existing patient-specific classifiers in terms of sensitivity-vs-specificity and Mathews correlation coefficients.**

***Index Terms*—ECG classification; Arrhythmias; Deep neural networks; i-vectors; DNN adaptation**

I. Introduction

Heart arrhythmias refer to the irregular heartbeats of patients. Arrhythmias can be detected through electrocardiography (ECG), which is a process of recording the electrical activities of the heart. The availability of personal portable devices opens up the possibility of continuous monitoring of the hearts of arrhythmia patients. However, continuous monitoring will lead to a large volume of ECG data that require automatic analysis and classification by machines. One of the big challenges in automatic heartbeat classification is the variations in ECG characteristics among different patients, which is known as inter-patient variations. Patient-independent classifiers that are trained on the ECG of a large number of patients may not perform well on unseen patients. To address this issue, we developed a patient-specific ECG classifier that is adaptive to the ECG characteristics of individual patients. The classifier adopts a beat-by-beat analysis strategy to detect some types of arrhythmias (i.e., supraventricular- and ventricular-ectopic beats) during long-term continuous heart monitoring.

The i-vector approach [1] was originally proposed for speaker verification. It converts variable-length utterances into low-dimensional fixed-length vectors that capture the information of speakers. In this work, we used i-vectors to represent the patient-specific characteristics in the ECG signals. For each patient, an i-vector is extracted from his/her ECG signals, which is then used for adapting a patient-independent DNN to make it tuned to the characteristics of the patient.

This work was in part supported by the RGC of Hong Kong, Grant No. PolyU152137/17E.

Recently, there have been much effort [2]–[5] in developing patient-specific heartbeat classifiers. These studies adopted a "subject-oriented" evaluation scheme [6] and used five minutes of patient-specific ECG data for constructing patient-specific classifiers. Under this scheme, data are divided according to patients instead of heartbeats, which ensures that the training set and the test set comprise different patients. As a result, the performance reported in [2]–[5] is more realistic and is closer to the practical situations.

In [2]–[4], the patient-specific classifiers were trained based on common and patient-specific beats. The common heartbeats were randomly sampled from the general population so that their number was limited to a few hundred only. In [5], instead of using the ECG of the entire population, a subset was selected for training the general classifier. However, reducing the amount of data from the general population is not a desirable way to address the data imbalance problem because it throws away lots of useful information in the ECG data. Also, the common training data are useful when the patient-specific beats contain a few arrhythmia patterns only [7].

In our adaptation method, all of the ECG data from the general population are used for training a DNN. Then, for each patient, an i-vector is extracted from his/her 5-minute ECG data. The i-vector is used as an input to the middle layer of the DNN and the whole network is fine-tuned by backpropagation. The advantage of the method is that it can leverage all of the ECG data in the general population but still be able to adapt to the ECG characteristics of individual patients through the i-vectors.

II. I-vector Adapted DNN

A. I-vector Extraction

The idea of i-vectors is based on the factor analysis method that compresses speaker and channel information into a low-dimensional subspace [8]. Inspired by the success of i-vectors in speaker recognition, we applied i-vectors to represent patient-specific information in ECG signals. Fig. 1 illustrates the procedure of training an i-vector extractor given a set of ECG data from a general population and the process of extracting an i-vector from an ECG record. First, PCA whitening is applied to reduce the correlation among the time-points in the ECG vectors [9]. Then, the whitened ECG vectors from the general population are used to train a Gaussian mixture model, which we referred to as the universal background model (UBM). The ECG data are then aligned with the UBM to compute the 0th- and 1st- order sufficient statistics,

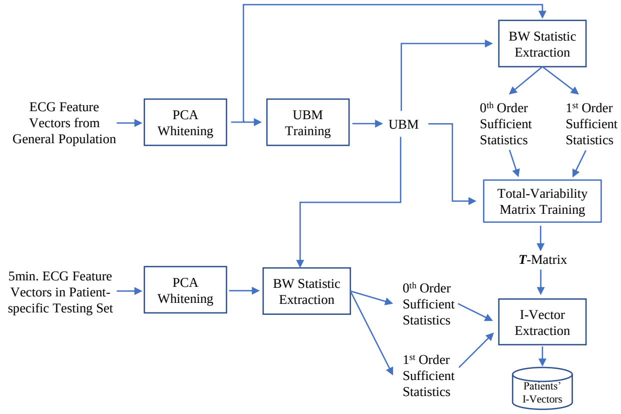

Fig. 1. Training of i-vector extractor and i-vector extraction process.

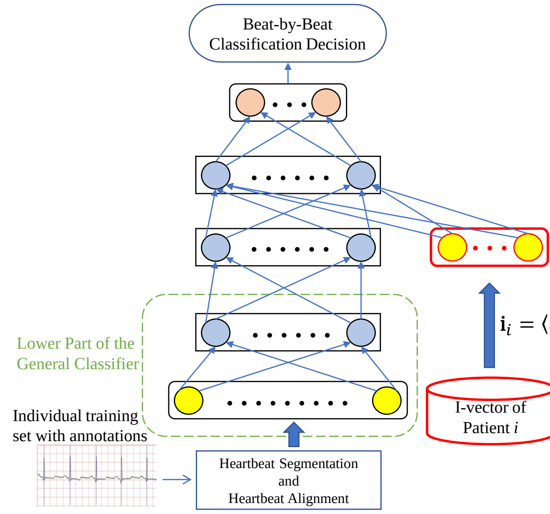

Fig. 2. I-vector adapted patient-specific DNN (iAP-DNN).

from which a total variability matrix (T-matrix) is trained. To extract an i-vector, the same processing pipeline is applied (see the lower branch of Fig. 1) to an ECG record to compute the sufficient statistics. Given the T-matrix and the sufficient statistics, an i-vector representing the whole ECG record can be obtained. In the sequel, we outline the formulae for training an i-vector extractor and the i-vector extraction process. For detailed derivations, readers may refer to [10].

Given the i-th ECG record from a general population, we extract the D-dimensional ECG vectors $\mathcal{X}_i = \{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT_i}\}$ from the record, where T_i is the number of complete heartbeats in the record.[1] We assume that the ECG vectors from this record are generated by a C-mixture GMM with parameters $\Lambda_i = \{\pi_c, \boldsymbol{\mu}_{ic}, \boldsymbol{\Sigma}_c\}_{c=1}^{C}$, i.e.,

$$p(\mathbf{x}_{it}) = \sum_{c=1}^{C} \pi_c^{(b)} \mathcal{N}(\mathbf{x}_{it} | \boldsymbol{\mu}_{ic}, \boldsymbol{\Sigma}_c^{(b)}), \qquad t = 1, \ldots, T_i. \quad (1)$$

In Eq. 1, we assume that $\pi_c^{(b)}$ and $\boldsymbol{\Sigma}_c^{(b)}$ are tied across all ECG records and are equal to the mixture weights and covariance matrices of the UBM, respectively.

In the i-vector framework [1], the mean vectors $\{\boldsymbol{\mu}_{ic}\}_{c=1}^{C}$ are stacked to form a GMM-supervector [8] $\boldsymbol{\mu}_i = [\boldsymbol{\mu}_{i1}^{\mathsf{T}} \ \ldots \ \boldsymbol{\mu}_{iC}^{\mathsf{T}}]^{\mathsf{T}}$, which is assumed to be generated by the following factor analysis model [11]:

$$\boldsymbol{\mu}_i = \boldsymbol{\mu}^{(b)} + \boldsymbol{T}\boldsymbol{w}_i, \quad (2)$$

where $\boldsymbol{\mu}^{(b)}$ is obtained by stacking the mean vectors of the UBM, $\boldsymbol{T}$ is a $CD \times R$ low-rank total variability matrix modeling all sort of variability in the ECG vectors, and $\boldsymbol{w}_i \in \Re^R$ comprises the latent (total) factors. Eq. 2 suggests that the generated supervectors $\boldsymbol{\mu}_i$'s have mean $\boldsymbol{\mu}^{(b)}$ and convariance matrix $\boldsymbol{T}\boldsymbol{T}^{\mathsf{T}}$. Eq. 2 can also be written in a component-wise form:

$$\boldsymbol{\mu}_{ic} = \boldsymbol{\mu}_c^{(b)} + \boldsymbol{T}_c \boldsymbol{w}_i, \qquad c = 1, \ldots, C \quad (3)$$

where $\boldsymbol{\mu}_{ic} \in \Re^D$ is the c-th sub-vector of $\boldsymbol{\mu}_i$ (similarly for $\boldsymbol{\mu}_c^{(b)}$) and $\boldsymbol{T}_c$ is an $D \times R$ sub-matrix of $\boldsymbol{T}$.

[1]See [9] for the definition of complete heartbeats and their extraction procedure.

In the i-vector framework, every ECG record is assumed to be obtained from a different patient. As a result, the ECG vectors of Record i aligning to mixture c have mean $\boldsymbol{\mu}_{ic}$ and covariance matrix $\boldsymbol{\Sigma}_c^{(b)}$. This matrix measures the deviation of the ECG vectors from $\boldsymbol{\mu}_{ic}$. In practice, $\boldsymbol{\mu}_c^{(b)}$ and $\boldsymbol{\Sigma}_c^{(b)}$ are the mean vectors and covariance matrices of the UBM. As a result, we only need to estimate the T-matrix $\boldsymbol{T}$ from a set of training ECG vectors.

Assume that there are P ECG recordings from the general population. The T-matrix can be estimated according to the expectation-maximization (EM) algorithm as follows [10]:

- E-step:

$$\langle \boldsymbol{w}_i | \mathcal{X}_i \rangle = \boldsymbol{L}_i^{-1} \sum_{c=1}^{C} \boldsymbol{T}_c^{\mathsf{T}} (\boldsymbol{\Sigma}_c^{(b)})^{-1} \tilde{\boldsymbol{f}}_{ic}, \quad (4)$$

$$\langle \boldsymbol{w}_i \boldsymbol{w}_i^{\mathsf{T}} | \mathcal{X}_i \rangle = \boldsymbol{L}_i^{-1} + \langle \boldsymbol{w}_i | \mathcal{X}_i \rangle \langle \boldsymbol{w}_i | \mathcal{X}_i \rangle^{\mathsf{T}}, \quad (5)$$

$$\boldsymbol{L}_i = \boldsymbol{I} + \sum_{c=1}^{C} N_{ic} \boldsymbol{T}_c^{\mathsf{T}} (\boldsymbol{\Sigma}_c^{(b)})^{-1} \boldsymbol{T}_c; \quad (6)$$

- M-step:

$$\boldsymbol{T}_c = \left[\sum\nolimits_i \tilde{\boldsymbol{f}}_{ic} \langle \boldsymbol{w}_i | \mathcal{X}_i \rangle^{\mathsf{T}} \right] \left[\sum\nolimits_i N_{ic} \langle \boldsymbol{w}_i \boldsymbol{w}_i^{\mathsf{T}} | \mathcal{X}_i \rangle \right]^{-1}, \quad (7)$$

where $i = 1, \ldots, P$, $\langle \cdot | \cdot \rangle$ is conditional expectation and $\boldsymbol{T}_c$ is the c-th partition of $\boldsymbol{T}$. The 0th-order and the 1st-order Baum-Welch statistics in Eq. 4, Eq. 6 and Eq. 7 can be computed as follows:

$$\begin{aligned} N_{ic} &= \sum\nolimits_t \gamma_c(\mathbf{x}_{it}), \\ \tilde{\boldsymbol{f}}_{ic} &= \sum\nolimits_t \gamma_c(\mathbf{x}_{it})(\mathbf{x}_{it} - \boldsymbol{\mu}_c^{(b)}), \end{aligned} \quad (8)$$

where $\gamma_c(\mathbf{x}_{it})$ is the posterior probability of mixture c. The i-vector $\mathbf{i}_i = \langle \boldsymbol{w}_i | \mathcal{X}_i \rangle$ representing the i-th patient can be computed according to Eq. 4.

TABLE I
PERFORMANCE OF THE PATIENT-SPECIFIC CLASSIFIERS IN [2]–[4] AND OUR IAP-DNN (EXP. 1)

Method		[2]	[3]	[4]	iAP-DNN
Class S	ACC	96.6	96.1	96.4	**98.8**
	SEN	50.6	62.1	64.6	**76.7**
	SPC	98.8	98.5	98.6	**99.7**
	PPV	67.9	56.7	62.1	**92.9**
Class V	ACC	98.1	97.6	98.6	**98.8**
	SEN	86.6	83.4	**95**	93.4
	SPC	99.3	98.1	98.1	**99.4**
	PPV	93.3	87.4	89.5	**94.6**

TABLE II
PERFORMANCE COMPARISON IN TERMS OF MCCS (EXP. 1)

Method		[2]	[3]	[4]	iAP-DNN
Class	N	0.83	0.81	0.84	**0.92**
	S	0.57	0.57	0.62	**0.84**
	V	0.87	0.83	0.91	**0.93**
	F	0.55	0.67	0.78	**0.82**
	Q	0	0	0	**0**
OMCC		0.93	0.92	0.94	**0.97**

B. Patient-Specific DNN

A patient-independent DNN (general classifier) is trained using the ECG data of a general population. This DNN receives segmented and aligned heartbeats [9] as input and heartbeat classes as output. To create a patient-specific classifier, the weights in the lower part of the general classifier are retained and the weights in the upper part are randomized. Then, for each patient, five minutes of his/her ECG data are presented to the input and an i-vector extracted from these 5-minute ECG data is presented to the middle layer of the patient-independent DNN, as shown in Fig. 2. The whole network is then fined tuned by backpropagation (BP).

The i-vector is presented to the second hidden layer instead of the first hidden layer because it is well known that the feature representation becomes increasingly abstract when moving up the network [12]. For example, in DNN-based speech recognition, the bottom layers can capture low-level acoustic features that vary significantly across different speakers and the upper layers can capture high-level features that are less speaker dependent [13]. This suggests that the upper layer can implicitly normalize the features across speakers. By the same token, the upper layers of the DNN in Fig. 2 will produce patient-invariant features, which is not good for patient-specific classification. To make the output patient-dependent, we inject patient-specific i-vectors to the middle layer of the DNN. The BP algorithm will encourage the upper layers to represent patient-dependent ECG information at a more abstract level. This results in the output layer being tuned to the characteristics of the corresponding patient.

III. EXPERIMENTAL SETTING

The MIT-BIH arrhythmia database [7] was used for performance evaluation. The database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings of 47 patients. Each record contains a continuous recording of raw ECG signals, which were digitized at 360 samples per second per channel with 11-bit resolution over a 10mV range. The database provides annotation for both beat-by-beat class information and corresponding time series information (e.g., positions of R peaks) that were verified by two or more cardiologists independently. The total number of labelled heartbeats is 108,655 and these heartbeats are classified into 15 different types. According to the AAMI recommendation [14], the 15 heartbeat types are classified into five classes, i.e., normal sinus beats (N), supraventricular ectopic beats (S), ventricular ectopic beats (V), fusion of a normal and a ventricular ectopic beat (F) and unknown beat type (Q). As suggested by the ANSI/AAMI EC57 standard [14], we focused on evaluating the classification performance of two majority arrhythmia classes (Classes S and V). We conducted two experiments (Exp. 1 and Exp. 2) to compare the performance of the iAP-DNN with five state-of-the-art patient-specific classifiers [2]–[5]. Under the same experimental protocol, the performance of different patient-specific classifiers can be compared directly.

We applied the aligned 417-dimensional feature vectors as described in [9] to train a general classifier. The general classifier has three hidden layers with a structure 417-100-100-100-5. We used the rectified linear unit (ReLU) in the hidden layers. The Adam optimizer [15] with default parameters was used for stochastic mini-batch (batch size of 128) gradient descent. Batch normalization and dropout were employed to train the DNNs. A dropout layer was added between the input and the first hidden layer, and the dropout rate was set to 20%. To train the i-vector extractor, we investigated different numbers of mixture components in the UBM (e.g., 16 and 20) and different i-vector dimensions (e.g., 32, 64 and 128), and the optimal combination was found to be 20 and 64 for the number of mixtures and i-vector dimension, respectively.

Same as [2]–[5], the classification performance on each heartbeat class was measured by using four standard metrics, namely, classification accuracy (ACC), sensitivity (SEN), specificity (SPC) and positive predictive value (PPV). Moreover, Matthews correlation coefficient (MCC) [16], [17] was calculated to measure the performance of different classifiers. MCC can reflect the performance of classifiers under serve data-imbalance scenarios.

In addition, receiver operating characteristics (ROCs) [18] were used to show the tradeoff between the performance measures (i.e., SEN vs. SPC) of a binary classifier when the decision threshold varies. Because the threshold typically has a wide range, ROC curves can provide more comprehensive information on performance.

IV. PERFORMANCE INVESTIGATION

A. Tests on 24 ECG recordings

The first experiment (Exp. 1) was conducted to evaluate the proposed methods based on 24 ECG recordings. The performance of iAP-DNN and that of [2]–[4] are shown in Table I. Except for the SEN of Class V in [4], the overall

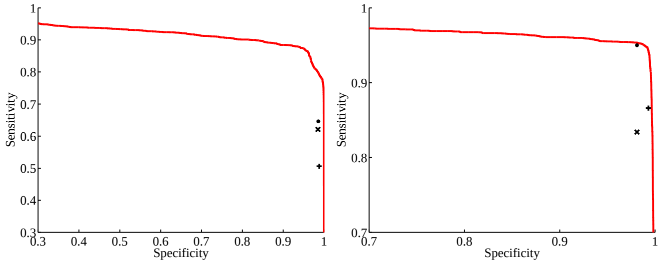

Fig. 3. ROC curves (SEN vs. SPC) of iAP-DNN (Exp. 1). *Left panel*: Class S vs. non S (AUC = 0.933). *Right panel*: Class V vs. non V (AUC = 0.972). Black markers correspond to the best performance in [2]–[4]. *AUC*: Area under the ROC curve [19].

TABLE III
PERFORMANCE OF THE PATIENT-SPECIFIC CLASSIFIERS IN [2], [3], [5] AND OUR IAP-DNN (EXP. 2)

Method		[2]	[3]	[5] Method I	[5] Method II	iAP-DNN
Class S	ACC	97.5	96.1	**99.1**	98.3	**99.1**
	SEN	74.9	**81.8**	76.5	61.4	78.8
	SPC	98.8	98.5	**99.9**	99.8	**99.9**
	PPV	78.8	63.4	**99.1**	90.7	98.7
Class V	ACC	98.8	97.9	**99.7**	99.4	**99.7**
	SEN	94.3	90.3	97.1	91.8	**97.4**
	SPC	99.4	98.8	**99.9**	**99.9**	**99.9**
	PPV	95.8	92.2	**98.5**	98.0	97.8

performance of the proposed method in Class S and Class V is significantly better than that in [2]–[4] for all evaluation measures.

Table II shows the performance in terms of MCCs between the proposed patient-specific classifiers and the patient-specific classifiers in [2]–[4]. Note that OMCC refers to overall MCC of the five classes. We can see that the MCC of the proposed method is much higher than the other three classifiers.

Fig. 3 shows the ROC curves of the proposed method in Class S and Class V. In Fig. 3, the operating points of the best performing classifiers in [2]–[4] are also shown as plus signs, crosses and points, respectively. The figures clearly show that the sensitivity-specificity points in [2]–[4] are below the red curve. This means that, within a certain range of decision thresholds, the iAP-DNN achieves better performance in term of both sensitivity and specificity than the classifiers in [2]–[4].

B. Tests on 22 ECG recordings

The second experiment (Exp. 2) was conducted to evaluate the proposed method based on 22 ECG recordings. Table III shows the ACC, SEN, SPC and PPV of the four methods in Class S and V. Note that in Method I [5], five minutes of labelled ECG signals of a patient was used to adapt the patient-specific classifier. Method II is the same as Method I but the manual labeling process is not required.

In Table III, for Class V, the SEN of the iAP-DNN is the highest among all methods and a high SPC (99.9%) is achieved. For Class S, although the SEN of the iAP-DNN is lower than that in [3], its SPC and PPV are higher.

V. CONCLUSION

We introduce an adaptive patient-specific heartbeat classification model for diagnosing heart arrhythmias, which leverages the DNNs for both feature extraction and classification based on the raw ECG signals. The general classifier was trained with all of the ECG recordings instead of selecting a subset. Then, the weights in the lower part of the general classifier were retained and the weights in the upper part were randomized to create a patient-specific classifier, and patient-dependent i-vectors were used for adaptation. The results show that the proposed iAP-DNN achieves better performance than existing patient-specific ECG classification systems.

REFERENCES

[1] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," IEEE Trans. Audio Speech Lang. Process., vol. 19, no. 4, pp. 788–798, 2011.

[2] W. Jiang and S. G. Kong, "Block-based neural networks for personalized ECG signal classification," IEEE Trans. Neural Net., vol. 18, no. 6, pp. 1750–1761, Nov. 2007.

[3] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ECG signals," IEEE Trans. Biomed. Eng., vol. 56, no. 5, pp. 1415–1426, 2009.

[4] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," IEEE Trans. Biomed. Eng., vol. 63, no. 3, pp. 664–675, Mar. 2016.

[5] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "An automatic subjectadaptable heartbeat classifier based on multiview learning," IEEE J. Biomed. Health Inform., vol. 20, no. 6, pp. 1482–1492, Nov. 2016.

[6] C. Ye, B. V. K. V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ECG signals," IEEE Trans. Biomed. Eng., vol. 59, no. 10, pp. 2930–2941, Oct. 2012.

[7] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," IEEE Eng. Med. Biol., vol. 20, no. 3, pp. 45–50, May 2001.

[8] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," IEEE Signal Process. Lett., vol. 13, no. 5, pp. 308–311, 2006.

[9] S. S. Xu, M. W. Mak, and C. C. Cheung, "Towards end-to-end ECG classification with raw signal extraction and deep neural networks," IEEE J. Biomed. Health Inform.: DOI 10.1109/JBHI.2018.2871510.

[10] M. W. Mak, "Lecture notes on factor analysis and i-vectors," Tech. Rep., Dept. of Electronic and Information Engineering, The Hong Kong Polytechnic University, 2016.

[11] C. M. Bishop, Pattern Recognition and Machine Learning. Springer-Verlag New York Inc., 2006.

[12] Y. Bengio, A. Courville, and P. Vincent,"Representation learning: A review and new perspectives," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, 2013.

[13] A. Mohamed, G.Hinton, and G. Penn, "Understanding how deep belief networks perform acoustic modelling," in IEEE Int. Conf. on Acoustics, Speech and Signal Process., 2012, pp. 4273–4276.

[14] Assoc. Adv. Med. Instrument., "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms," ANSI/AAMI, no. EC57, 1998.

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[16] B. W. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme," Biochem. et Biophys. Acta (BBA)-Protein Struct., vol. 405, no. 2, pp. 442–451, 1985.

[17] M. W. Mak, J. Guo, and S. Y. Kung, "PairProSVM: protein subcellular localization based on local pairwise profile alignment and SVM," IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 5, no. 3, pp. 416–422, 2008.

[18] J. A. Swets, Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

[19] T. Fawcett, "An introduction to roc analysis," Pattern Recognit. Lett.,, vol. 27, no. 8, pp. 861–874, 2006.