2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)

BioExpDNN: Bioinformatic Explainable Deep Neural Network

Hao Fang
College of Mathematics and Computer Science
Fuzhou University
Fuzhou, China
hfang7@163.com

Cheng Shi
School of Computer Science and Engineering
Xi'an University of Technology
Xi'an, China
chengc_s@163.com

Chi-Hua Chen
College of Mathematics and Computer Science
Fuzhou University
Fuzhou, China
chihua0826@gmail.com

2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 978-1-7281-6215-7/20/$31.00 ©2020 IEEE DOI: 10.1109/BIBM49941.2020.9313113

***Abstract*—In recent years, machine learning is applied in the bioinformatics and medical fields to analyze relationships among biological features and behaviors. However, it is difficult to discover the significant features of large-scale datasets. A novel feature extraction method called bioinformatic explainable deep neural network (BioExpDNN) is proposed to filter the critical features with strong influences on the dataset and to explain the interaction of features. In the practical experiments, this study adopted three biomedical science datasets from the UCI (University of California, Irvine) Machine Learning Repository: (1). Cryotherapy Data Set (CDS) contains 6 attributes and 2 classes (i.e., recovery and non-recovery); (2). Cervical Cancer Behavior Risk Data Set (CCBRDS) consists of 18 attributes and 2 classes (i.e., cervical cancer patient and healthy body); (3). Heart Failure Clinical Records Data Set (HFCRDS) includes 12 clinical attributes and 2 classes (i.e., death and life). In comparison results, extracted features were considered as inputs of a classifier based on deep neural network for classification. The classification accuracy was selected as an evaluation factor to evaluate the performance of feature extraction methods. The experimental results showed that the classification accuracies of CDS, CCBRDS, and HFCRDS were 92.59%, 100%, and 78.9%, respectively.**

Keywords—bioinformatics explainable deep neural network, feature extraction, principal component analysis, Pearson correlation coefficient analysis

I. INTRODUCTION

Machine learning is critical to modern bioinformatics. Since machine learning technology can make full use of a large number of biological information data by extracting features of biological information and mining potential valuable information. Especially in the medical field, it is possible to find out the specific causes of the corresponding symptoms by studying the clinical features of a large number of data related to patients.

For feature extraction, some existed methods which include principal component analysis (PCA) [1]-[10], Pearson correlation coefficient analysis (PCCA) [11]-[14], and entropy analysis (EA) [13]-[26] have been proposed to estimate the important level of each attribute. Pearson correlation coefficient denotes the degree of similarity between each-two variables, and the information gain based on entropy denotes the important level of each attribute for the outputs of model. In the machine learning process, PCCA and EA only estimate the correlations between each attribute and each output, but the effects of interactions among features have not been analyzed by PCCA and EA. Although PCA can integrate the whole dataset to discover the critical features, the principal components cannot explain the implication of the bioinformatic data.

the optimal features for classification. The proposed BioExpDNN improves the explanatory power of deep neural networks to discover the interaction and significance of features. To extract the optimal features, the proposed BioExpDNN adds an explainable layer between the input layer and the first hidden layer to generate the weight of each input. The explainable layer is not fully connected to the input layer. Furthermore, a neuron in the input layer is connected with a neuron in the explainable layer. The weight of the connection between a neuron in the input layer (i.e., an attribute in a data set) and a neuron in the explainable layer denotes the weight of the attribute. By feeding the whole dataset into the BioExpDNN, a weight list can be generated and used to measure the significance of features. According to the weight list, the optimal features with higher weights can be extracted.

In the performance evaluation stage, the extracted features are regarded as the inputs of a classifier based on deep neural network for classification. This study takes practical experiments on three life science data sets from the UCI (University of California, Irvine) Machine Learning Repository which include (1). Cryotherapy Data Set (CDS) [27]-[29], (2). Cervical Cancer Behavior Risk Data Set (CCBRDS) [30]-[31], and (3). Heart Failure Clinical Records Data Set (HFCRDS) [32]-[33]. The experiments compare the classification accuracies of PCA, PCCA, and the proposed BioExpDNN.

The main contributions of this paper are listed as follows.

- For the case of CDS, the proposed BioExpDNN was used to extract 3 critical features (i.e., the surface area of warts, the types of warts, and the time elapsed before treatment) for estimating recovery.
- For the case of CCBRDS, the proposed BioExpDNN was used to extract 13 critical features (i.e., perception severity, empowerment knowledge, eating behavior, attitude consistency, social support instrument, social support appreciation, empowerment desires, motivation willingness, intention aggregation, perception vulnerability, intention commitment, personal hygiene behavior, and norm fulfillment) for detecting cervical cancer.
- For the case of HFCRDS, the proposed BioExpDNN was used to extract 11 critical features (i.e., creatinine phosphokinase, serum creatinine, serum sodium, time, non-high blood pressure, ejection fraction, platelets, smoking, age, non-anaemia, and high blood pressure) for predicting life status.

The rest of this paper is organized as follows. Section II introduces the structure of the proposed BioExpDNN and the

Authorized licensed use limited to: Birkbeck University of London. Downloaded on March 28,2021 at 03:35:31 UTC from IEEE Xplore. Restrictions apply.

experimental results. Section IV draws the conclusions of this work and discusses future work.

II. Feature Extraction Based on BioExpDNN

This section proposes the structure of BioExpDNN and give a case study to explain BioExpDNN.

A. The Structure of BioExpDNN

The traditional deep neural network (DNN) contains an input layer, several hidden layers, and an output layer (shown in Fig. 1). Furthermore, all layers in the traditional DNN are fully-connected layers, and each neuron includes corresponding weights and biases. For improving the explanatory power of the traditional DNN to extract critical features of bioinformatics, a bioinformatic explainable deep neural network (BioExpDNN) is proposed to add an explainable layer [34] between the input layer and the first hidden layer (shown in Fig. 2).

For instance, n neurons will be built in the explainable layer without fully-connection if the input layer includes n inputs (i.e., an input set $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$). The i-th input connects to the i-th neuron in the explainable layer. The activation function of each neuron in the explainable layer is a linear function with a weight (w_i), and the activation level of the i-th neuron in the explainable layer (ψ_i) is calculated by Equation (1). Furthermore, the proposed BioExpDNN includes q hidden layers with non-linear activation functions to analyze non-linear relationships, and the notation of a non-linear activation function is defined as $\varphi(\bullet)$. Equation (2) denotes the activation level of the j-th neuron ($h_j^{(1)}$) with a bias ($b_j^{(1)}$) in the first hidden layer, and the weight of connection between the r-th neuron in the explainable layer and the j-th neuron in the first hidden layer is $w_{r,j}^{(1)}$; Equation (3) denotes the activation level of the j-th neuron ($h_j^{(v)}$) with a bias ($b_j^{(v)}$) in the v-th hidden layer, and the weight of connection between the r-th neuron in the (v-1)-th layer and the j-th neuron in the v-th hidden layer is $w_{r,j}^{(v)}$; the number of neurons in the v-th hidden layer is k_v. For classification, the function $\vartheta(\bullet)$ denotes a softmax function in the output layer, and the weight of connection between the r-th neuron in the q-th layer and the j-th neuron in the output layer is $w_{r,j}^{(o)}$. The output layer of BioExpDNN includes m outputs (i.e., an output set $\mathbf{Y} = \{y_1, y_2, \ldots, y_m\}$), and Equation (4) denotes the value of the j-th output with a bias ($b_j^{(o)}$).

$$\psi_i = w_i \times x_i . \tag{1}$$

$$h_j^{(1)} = \varphi\left(\sum_{r=1}^{n} w_{r,j}^{(1)} \times \psi_r + b_j^{(1)} \right). \tag{2}$$

$$h_j^{(v)} = \varphi\left(\sum_{r=1}^{k_{v-1}} w_{r,j}^{(v)} \times h_r^{(v-1)} + b_j^{(v)} \right). \tag{3}$$

$$y_j = \vartheta\left(\sum_{r=1}^{k_q} w_{r,j}^{(o)} \times h_r^{(q)} + b_j^{(o)} \right). \tag{4}$$

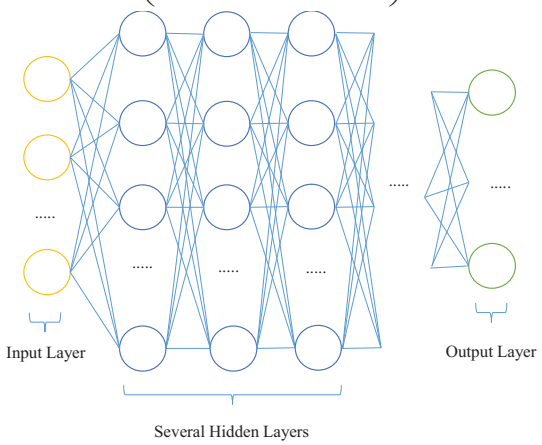

Fig. 1. The structure of DNN

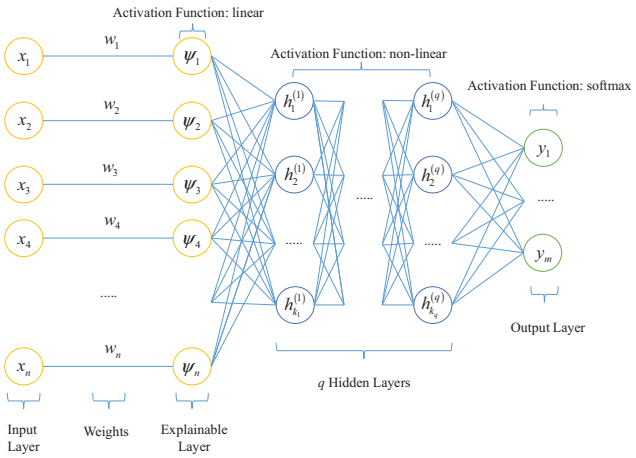

Fig. 2. The structure of BioExpDNN

For feature extraction, the weight list (i.e., the weight set $\mathbf{W} = \{w_1, w_2, \ldots, w_n\}$) can be obtained after training the BioExpDNN. The u critical features can be extracted with the top u absolute weights.

B. A Case Study of BioExpDNN

In this subsection, the cryotherapy data set (CDS) is selected as a case study to explain the proposed BioExpDNN. The attributes and classes in the CDS are showed in TABLE I, and the structure of BioExpDNN is illustrated in TABLE II and Fig. 3, and the detail description is presented in the following paragraphs.

TABLE I. The Attributes in the CDS

Notation	*Attribute*
a_1	The gender of patient
a_2	The age of patient
a_3	The time elapsed before treatment
a_4	The number of warts on a patient
a_5	The types of warts
a_6	The surface area of warts
c_1	Recovery
c_2	Non-recovery

Authorized licensed use limited to: Birkbeck University of London. Downloaded on March 28,2021 at 03:35:31 UTC from IEEE Xplore. Restrictions apply.

TABLE II. THE STRUCTURE OF BIOEXPDNN FOR THE CDS

Layer	*The Number of Neurons*	*Activation Function*	*Having Biases?*
Input Layer	6	none	No
Explainable Layer	**6**	**linear**	**No**
Hidden Layer 1	4	tanh	Yes
Hidden Layer 2	4	tanh	Yes
Hidden Layer 3	4	tanh	Yes
Output Layer	2	softmax	Yes

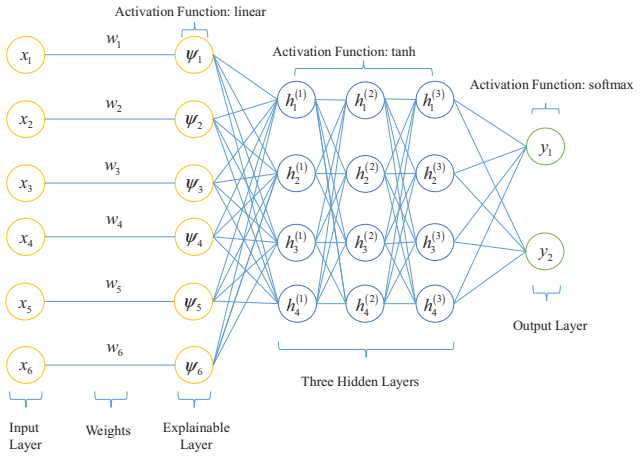

Fig. 3. The structure of BioExpDNN for the CDS

- For setting the inputs of BioExpDNN, the CDS consists of 6 attributes (i.e., an attribute set $\mathbf{A} = \{a_1, a_2, a_3, a_4, a_5, a_6\}$) which include (1). the gender of patient, (2). the age of patient, (3). the time elapsed before treatment, (4). the number of warts on a patient, (5). the types of warts, and (6). the surface area of warts. The attribute set **A** was adopted as the input set **X** of BioExpDNN (i.e., $\mathbf{X} = \{x_1, x_2, x_3, x_4, x_5, x_6\}$), so the value of n was 6 in this case.
- For setting the hidden layers of BioExpDNN, three hidden layers (i.e., the value of q was equal to 3) were built, and the tanh function was adopted as the activation function for the neurons of these three hidden layers. The number of neurons in each hidden layer was 4, so the numbers of k_1, k_2, and k_3 were equal to 4, respectively. Equation (5) denotes the activation level of the j-th neuron ($h_j^{(v)}$) in the v-th hidden layer based on the tanh function.

$$h_j^{(v)} = \varphi\left(\sum_{r=1}^{4} w_{r,j}^{(v)} \times h_r^{(v-1)} + b_j^{(v)}\right), \quad \text{where } \varphi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}. \tag{5}$$

- For setting the outputs of BioExpDNN, the CDS contains 2 classes (i.e., a class set $\mathbf{C} = \{c_1, c_2\}$) which include (1). recovery and (2). non-recovery. The class set **C** was adopted as the output set **Y** of BioExpDNN (i.e., $\mathbf{Y} = \{y_1, y_2\}$), so the value of m was 2 in this case.
- For training the BioExpDNN, the adaptive moment estimation method (ADAM) was adopted as the optimizer to update the weights in the BioExpDNN.

After training the BioExpDNN, the value of each weight in the explainable layer (i.e., the weight set $\mathbf{W} = \{w_1, w_2, w_3, w_4, w_5, w_6,\}$) was obtained. In this case study, the absolute values of these weights (i.e., $|w_1|$, $|w_2|$, $|w_3|$, $|w_4|$, $|w_5|$, and $|w_6|$)(shown in TABLE III) were determined as 0.89, 1.73, 2.27, 1.29, 2.83, and 3.76, respectively. The important levels of attributes can be ordered as (1). a_6: the surface area of warts, (2). a_5: the types of warts, (3). a_3: the time elapsed before treatment, (4). a_2: the age of patient, (5). a_4: the number of warts on a patient, and (6). a_1: the gender of patient. Therefore, the 3 critical features (i.e., (1). a_6: the surface area of warts, (2). a5: the types of warts, and (3). a_3: the time elapsed before treatment) can be extracted with the top 3 absolute weights (i.e., the value of u was equal to 3). These three features are extracted and adopted as the input of a classifier for estimating recovery.

TABLE III. THE WEIGHT FOR EACH ATTRIBUTE IN THE CRYOTHERAPY DATA SET

Attribute Notation	*Attribute*	*Weight Value*
a_1	The gender of patient	0.89
a_2	The age of patient	1.73
a_3	The time elapsed before treatment	2.27
a_4	The number of warts on a patient	1.29
a_5	The types of warts	2.83
a_6	The surface area of warts	3.76

III. CASE STUDIES AND EXPERIMENTAL RESULTS

This section describes the experimental environments and illustrates the practical experimental results of three biomedical datasets for the evaluation of BioExpDNN.

A. Experimental Environments

In experiments, three life science datasets from the UCI Machine Learning Repository including (1). CDS, (2). CCBRDS, and (3). HFCRDS were selected to perform the two-classes classification task. Furthermore, three approaches (i.e., PCA, PCCA, and BioExpDNN) were considered to filter critical features for the comparison of feature extraction ability. The extracted features by these methods were adopted as the inputs of a DNN-based classifier for classification. The structure of the DNN-based classifier included three hidden layers, and each hidden layer had four neurons in (shown in TABLE IV and Fig. 4). For the evaluation of feature extraction, the classification accuracy of each method was measured. The training set contains 70% of a dataset, and the testing set contains the other 30% of the dataset; The training set and the testing set were selected by stratified random sampling.

B. Case (1): Cryotherapy Data Set

In the case of CDS, the dataset consists of 6 attributes (i.e., (1). the gender of patient, (2). the age of patient, (3). the time elapsed before treatment, (4). the number of warts on a patient, (5). the types of warts, and (6). the surface area of warts) which are described in Subsection II. B. The weight of each attribute by the three feature extraction methods is showed in TABLE V.

Authorized licensed use limited to: Birkbeck University of London. Downloaded on March 28,2021 at 03:35:31 UTC from IEEE Xplore. Restrictions apply.

TABLE IV. THE STRUCTURE OF BIOEXPDNN FOR THE CDS

Layer	*The Number of Neurons*	*Activation Function*	*Having Biases?*
Input Layer	6	none	No
Hidden Layer 1	4	tanh	Yes
Hidden Layer 2	4	tanh	Yes
Hidden Layer 3	4	tanh	Yes
Output Layer	2	softmax	Yes

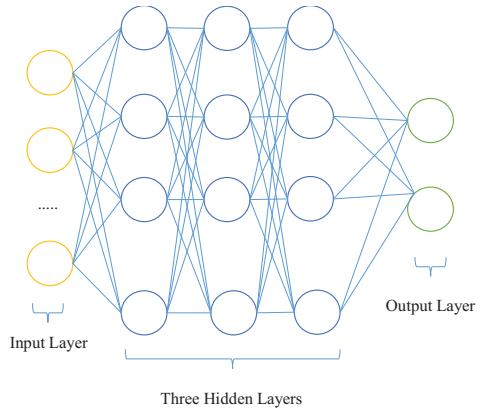

Fig. 4. The structure of a DNN-based classifier for classification

TABLE V. THE ABSOLUTE WEIGHT FOR EACH ATTRIBUTE IN THE CRYOTHERAPY DATA SET (CDS)

Notation	*Attribute*	*PCA*	*PCCA*	*BioExpDNN*
a_1	The gender of patient	0.19	0.09	0.89
a_2	The age of patient	0.45	0.54	1.73
a_3	The time elapsed before treatment	0.45	**0.65**	2.27
a_4	The number of warts on a patient	0.01	0.08	1.29
a_5	The types of warts	**0.60**	0.49	2.83
a_6	The surface area of warts	0.45	0.19	**3.76**

The results of PCA showed the absolute values of weights were determined as 0.19, 0.45, 0.45, 0.01, 0.60, and 0.45, respectively. Therefore, the important levels of attributes were ordered as a_5, a_6, a_2, a_3, a_1, and a_4. In this case, the most important extracted feature was a_5 (i.e., the types of warts) by PCA. However, PCA only considered the relationships among inputs to find the features which were fitting the trend of inputs. Therefore, the feature of a_5 may indicate the trend of the attribute set **A**, but the relationships between inputs and outputs were not considered and analyzed by PCA.

The results of PCCA showed the absolute values of weights were determined as 0.09, 0.54, 0.65, 0.08, 0.49, and 0.19, respectively. Therefore, the important levels of attributes were ordered as a_3, a_2, a_5, a_6, a_1, and a_4. In this case, the most important extracted feature was a_3 (i.e., the time elapsed before treatment) by PCCA. Although the feature a_3 with the higher correlation coefficient between inputs and outputs was extracted, the inputs were assumed as independent variables. Therefore, the interactions of features were not considered and analyzed by PCCA.

The results of BioExpDNN showed the absolute values of weights were determined as 0.89, 1.73, 2.27, 1.29, 2.83, and 3.76, respectively. Therefore, the important levels of attributes were ordered as a_6, a_5, a_3, a_2, a_4, and a_1. The critical features a_5 (i.e., the most import feature by PCA) and a_3 (i.e., the most import feature by PCCA) were also extracted; the unimportant features a_1 and a_4 were also filtered out. Furthermore, the proposed BioExpDNN can analyze the trend of inputs and the relationships between inputs and outputs; the combination of a_6, a_5, and a_3 was more suitable for estimating the outputs (i.e., the classification in this case).

For classification and evaluation, the top 3 critical features were selected to be adopted as the inputs of the DNN-based classifier (shown in Fig. 4). In the case of CDS, three classifiers (i.e., Classifier 1, Classifier 2, and Classifier 3) were trained by the training set and tested by the testing set. The testing results of each classifier were presented as follows.

- Classifier 1 considered the extracted features a_5, a_6, and a_2 from PCA, and the accuracy of Classifier 1 was 85.19%.
- Classifier 2 considered the extracted features a_3, a_2, and a_5 from PCCA, and the accuracy of Classifier 2 was 81.48%.
- Classifier 3 considered the extracted features a_6, a_5, and a_3 from BioExpDNN, and the accuracy of Classifier 3 was 92.59%.

C. Case (2): Cervical Cancer Behavior Risk Data Set

In the case of CCBRDS, the dataset consists of 18 attributes (i.e., an attribute set $\mathbf{D} = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}, d_{11}, d_{12}, d_{13}, d_{14}, d_{15}, d_{16}, d_{17}, d_{18}\}$) which include (1). Patient's eating habits, (2). Patient's personal hygiene habits, (3). the determinant of intention behavior, (4). the determinant of intention behavior, (5). the degree of attitude consistency, (6) the degree of attitude spontaneity, (7). the significance of a person towards society, (8). the degree of fulfillment, (9). the degree of perceived vulnerability, (10). the degree of perceived severity, (11). motivation strength, (12). motivation willingness, (13). the influence of emotions on behavior, (14). the influence of appreciation on behavior, (15). the influence of instrumental on behavior, (16). the influence of knowledge on behavior, (17). the influence of abilities on behavior, and (18). the influence of desires on behavior (shown in TABLE VI). The attribute set **D** was adopted as the input set **X** of BioExpDNN (i.e., $\mathbf{X} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}\}$), so the value of n was 18 in this case. Furthermore, the CCBRDS contains 2 classes (i.e., a class set $\mathbf{G} = \{g_1, g_2\}$) which include (1). cervical cancer patient and (2). healthy body. The class set **G** was adopted as the output set **Y** of BioExpDNN (i.e., $\mathbf{Y} = \{y_1, y_2\}$), so the value of m was 2 in this case.

The results of PCA showed the absolute values of weights were determined as 0.01, 0.22, 0.13, 0.06, 0.03, 0.01, 0.02, 0.01, 0.05, 0.05, 0.20, 0.34, 0.37, 0.35, 0.32, 0.37, 0.38, and 0.35, respectively. Therefore, the important levels of attributes were ordered as d_{17}, d_{16}, d_{13}, d_{18}, d_{14}, d_{12}, d_{15}, d_2, d_{11}, d_3, d_4, d_{10}, d_9, d_5, d_7, d_6, d_8, and d_1. In this case, the most important extracted feature was d_{17} (i.e., the influence of abilities on behavior) by PCA. Furthermore, The results of

Authorized licensed use limited to: Birkbeck University of London. Downloaded on March 28,2021 at 03:35:31 UTC from IEEE Xplore. Restrictions apply.

PCCA showed the absolute values of weights were determined as 0.20, 0.37, 0.29, 0.24, 0.13, 0.08, 0.29, 0.43, 0.42, 0.51, 0.47, 0.43, 0.39, 0.31, 0.13, 0.48, 0.54, and 0.46, respectively. Therefore, the important levels of attributes were ordered as d_{17}, d_{10}, d_{16}, d_{11}, d_{18}, d_{12}, d_8, d_9, d_{13}, d_2, d_{14}, d_7, d_3, d_4, d_1, d_{15}, d_5, and d_6. In this case, the most important extracted feature was d_{17} (i.e., the influence of abilities on behavior) by PCCA. Therefore, the feature d_{17} (i.e., the influence of abilities on behavior) was extracted in accordance with the relationships among inputs and the relationships between inputs and outputs.

TABLE VI. THE ABSOLUTE WEIGHT FOR EACH ATTRIBUTE IN THE CERVICAL CANCER BEHAVIOR RISK DATA SET (CCBRDS)

Notation	*Attribute*	*PCA*	*PCCA*	*BioExpDNN*
d_1	Patient's eating habits	0.01	0.20	1.56
d_2	Patient's personal hygiene habits	0.22	0.37	1.18
d_3	The determinant of intention behavior	0.13	0.29	1.30
d_4	The determinant of intention behavior	0.06	0.24	1.22
d_5	The degree of attitude consistency	0.03	0.13	1.49
d_6	The degree of attitude spontaneity	0.01	0.08	0.94
d_7	The significance of a person towards society	0.02	0.29	0.99
d_8	The degree of fulfillment	0.01	0.43	1.16
d_9	The degree of perceived vulnerability	0.05	0.42	1.21
d_{10}	The degree of perceived severity	0.05	0.51	**1.69**
d_{11}	Motivation strength	0.20	0.47	1.25
d_{12}	Motivation willingness	0.34	0.43	1.34
d_{13}	The influence of emotions on behavior	0.37	0.39	1.01
d_{14}	The influence of appreciation on behavior	0.35	0.31	1.45
d_{15}	The influence of instrumental on behavior	0.32	0.13	1.45
d_{16}	The influence of knowledge on behavior	0.37	0.48	1.60
d_{17}	The influence of abilities on behavior	**0.38**	**0.54**	1.16
d_{18}	The influence of desires on behavior	0.35	0.46	1.46

The results of BioExpDNN showed the absolute values of weights were determined as 1.56, 1.18, 1.3, 1.22, 1.49, 0.94, 0.99, 1.16, 1.21, 1.69, 1.25, 1.34, 1.01, 1.45, 1.45, 1.60, 1.16, and 1.46, respectively. Therefore, the important levels of attributes were ordered as d_{10}, d_{16}, d_1, d_5, d_{18}, d_{15}, d_{14}, d_{12}, d_3, d_{11}, d_4, d_9, d_2, d_8, d_{17}, d_{13}, d_7, and d_6. Although the feature d_{17} was extracted as a critical feature by PCA and PCCA, PCA and PCCA did not consider the interaction of features. The proposed BioExpDNN can analyze the trend of inputs and the relationships between inputs and outputs; the combination of d_{10}, d_{16}, d_1, d_5, d_{18}, d_{15}, d_{14}, d_{12}, d_3, d_{11}, d_4, d_9, and d_2 was more suitable for estimating the outputs (i.e., the classification in this case).

For classification and evaluation, the top 13 critical features were selected to be adopted as the inputs of the DNN-based classifier (shown in Fig. 4). In the case of CCBRDS, three classifiers (i.e., Classifier 4, Classifier 5, and Classifier 6) were trained by the training set and tested by the testing set. The testing results of each classifier were presented as follows.

- Classifier 4 considered the extracted features d_{17}, d_{16}, d_{13}, d_{18}, d_{14}, d_{12}, d_{15}, d_2, d_{11}, d_3, d_4, d_{10}, and d_9 from PCA, and the accuracy of Classifier 4 was 90.91%.
- Classifier 5 considered the extracted features d_{17}, d_{10}, d_{16}, d_{11}, d_{18}, d_{12}, d_8, d_9, d_{13}, d_2, d_{14}, d_7, and d_3 from PCCA, and the accuracy of Classifier 5 was 77.27%.
- Classifier 6 considered the extracted features d_{10}, d_{16}, d_1, d_5, d_{18}, d_{15}, d_{14}, d_{12}, d_3, d_{11}, d_4, d_9, and d_2 from BioExpDNN, and the accuracy of Classifier 6 was 100%. The results proved that the combination of d_{10}, d_{16}, d_1, d_5, d_{18}, d_{15}, d_{14}, d_{12}, d_3, d_{11}, d_4, d_9, and d_2 could obtain higher performance than the feature d_{17} which was extracted by PCA and PCCA.

D. Case (3): Heart Failure Clinical Records Data Set

In the case of HFCRDS, the dataset consists of 17 attributes (i.e., an attribute set $\mathbf{V} = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}, v_{14}, v_{15}, v_{16}, v_{17}\}$) which include (1). Patient's age, (2). anaemia, (3). no anemia, (4). creatinine phosphokinase (CPK)(mcg/L), (5). the symptoms of diabetes, (6). no symptoms of diabetes, (7). the percentage of blood leaving the heart at each contraction, (8). hypertension, (9). no hypertension, (10). the number of platelets in the blood (kiloplatelets/mL), (11). the level of serum creatinine in the blood (mg/dL), (12). the level of serum sodium in the blood (mEq/L), (13). male patient, (14). female patient, (15). smoking, (16). no smoking, and (17). follow-up period (days) (shown in TABLE VII). The attribute set $\mathbf{V}$ was adopted as the input set $\mathbf{X}$ of BioExpDNN (i.e., $\mathbf{X} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}\}$), so the value of n was 17 in this case. Furthermore, the HFCRDS contains 2 classes (i.e., a class set $\mathbf{L} = \{l_1, l_2\}$) which include (1). death and (2). life. The class set $\mathbf{L}$ was adopted as the output set $\mathbf{Y}$ of BioExpDNN (i.e., $\mathbf{Y} = \{y_1, y_2\}$), so the value of m was 2 in this case.

The results of PCA showed the absolute values of weights were determined as 0.02, 0.16, 0.16, 0.07, 0.22, 0.22, 0.10, 0.13, 0.13, 0.06, 0.00, 0.01, 0.45, 0.45, 0.44, 0.44, and 0.02, respectively. Therefore, the important levels of attributes were ordered as v_{13}, v_{14}, v_{15}, v_{16}, v_5, v_6, v_2, v_3, v_8, v_9, v_7, v_4, v_{10}, v_1, v_{17}, v_{12}, and v_{11}. In this case, the most important extracted feature was gender (i.e., v_{13}: male patient and v_{14}: female patient) by PCA. Furthermore, the results of PCCA showed the absolute values of weights were determined as 0.25, 0.07, 0.07, 0.06, 0.00, 0.00, 0.27, 0.08, 0.08, 0.05, 0.29, 0.20, 0.00, 0.00, 0.01, 0.01, and 0.53, respectively. Therefore, the important levels of attributes were ordered as v_{17}, v_{11}, v_7, v_1, v_{12}, v_8, v_9, v_2, v_3, v_4, v_{10}, v_{15}, v_{16}, v_{13}, v_{14}, v_5, and v_6. In this case, the most important extracted feature was v_{17} (i.e., follow-up period) by PCCA. Although the gender was extracted as a critical feature in accordance with the trend of inputs by PCA, PCA did not analyze the relationships between inputs and outputs. The results of PCCA indicated that the gender might be not a suitable feature for estimating the outputs.

Authorized licensed use limited to: Birkbeck University of London. Downloaded on March 28,2021 at 03:35:31 UTC from IEEE Xplore. Restrictions apply.

The results of BioExpDNN showed the absolute values of weights were determined as 2.83, 0.71, 2.40, 5.21, 0.73, 1.46, 3.59, 2.33, 3.59, 3.55, 4.45, 4.23, 1.07, 1.17, 2.84, 2.03, and 3.75, respectively. Therefore, the important levels of attributes were ordered as v_4, v_{11}, v_{12}, v_{17}, v_9, v_7, v_{10}, v_{15}, v_1, v_3, v_8, v_{16}, v_6, v_{14}, v_{13}, v_5, and v_2. The proposed BioExpDNN can analyze the interaction of inputs and the relationships between inputs and outputs; the combination of v_4, v_{11}, v_{12}, v_{17}, v_9, v_7, v_{10}, v_{15}, v_1, v_3, and v_8 was more suitable for estimating the outputs (i.e., the classification in this case).

TABLE VII. THE ABSOLUTE WEIGHT FOR EACH ATTRIBUTE IN THE HEART FAILURE CLINICAL RECORDS DATA SET (HFCRDS)

Notation	*Attribute*	*PCA*	*PCCA*	*BioExpDNN*
v_1	Patient's age	0.02	0.25	2.83
v_2	Anaemia	0.16	0.07	0.71
v_3	No anemia	0.16	0.07	2.40
v_4	Creatinine phosphokinase (CPK)(mcg/L)	0.07	0.06	**5.21**
v_5	The symptoms of diabetes	0.22	0.00	0.73
v_6	No symptoms of diabetes	0.22	0.00	1.46
v_7	The percentage of blood leaving the heart at each contraction	0.10	0.27	3.59
v_8	Hypertension	0.13	0.08	2.33
v_9	No hypertension	0.13	0.08	3.59
v_{10}	The number of platelets in the blood (kiloplatelets/mL)	0.06	0.05	3.55
v_{11}	The level of serum creatinine in the blood (mg/dL)	0.00	0.29	4.45
v_{12}	The level of serum sodium in the blood (mEq/L)	0.01	0.20	4.23
v_{13}	Male patient	**0.45**	0.00	1.07
v_{14}	Female patient	**0.45**	0.00	1.17
v_{15}	Smoking	0.44	0.01	2.84
v_{16}	No smoking	0.44	0.01	2.03
v_{17}	Follow-up period (days)	0.02	**0.53**	3.75

For classification and evaluation, the top 11 critical features were selected to be adopted as the inputs of the DNN-based classifier (shown in Fig. 4). In the case of HFCRDS, three classifiers (i.e., Classifier 7, Classifier 8, and Classifier 9) were trained by the training set and tested by the testing set. The testing results of each classifier were presented as follows.

- Classifier 7 considered the extracted features a_5, a_6, and a_2 from PCA, and the accuracy of Classifier 7 was 68.89%.
- Classifier 8 considered the extracted features a_3, a_2, and a_5 from PCCA, and the accuracy of Classifier 8 was 73.33%.
- Classifier 9 considered the extracted features v_4, v_{11}, v_{12}, v_{17}, v_9, v_7, v_{10}, v_{15}, v_1, v_3, and v_8 from BioExpDNN, and the accuracy of Classifier 9 was 78.89%.

E. Summary and Discussions

The classification accuracies of these three cases were summarized in TABLE VIII and Fig. 5. The results indicated the performance of the classifier with the proposed BioExpDNN in each case was higher than other classifiers with different feature extraction methods. Therefore, the proposed BioExpDNN is a powerful tool to find critical features and improve the explanatory power of DNNs.

TABLE VIII. THE COMPARISON OF CLASSIFICATION ACCURACIES

Dataset	*Classifier with PCA*	*Classifier with PCCA*	*Classifier with BioExpDNN*
CDS	85.19%	81.48%	**92.59%**
CCBRDS	90.91%	77.27%	**100%**
HFCRDS	68.89%	73.33%	**78.89%**

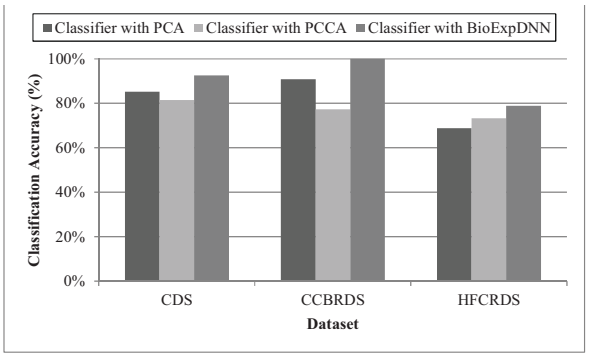

Fig. 5. The comparison of classification accuracies

IV. CONCLUSIONS AND FUTURE WORK

In this study, a BioExpDNN was proposed to filter the optimal features of bioinformatic datasets and to mine the potential beneficial information. The optimal features had huge impacts on the target states. In accordance with the extracted characteristics of large-scale bioinformatics data based on BioExpDNN, better classification results could be achieved. The main contributions of BioExpDNN were to draw the optimal features and to discover the interconnectedness among features. Furthermore, BioExpDNN utilized the interpretability of deep neural networks to explain the importance level of each feature based on the weights of the explainable layer. In the practical experimental results, the proposed method outperformed the other two approaches (i.e. PCA and PCCA). The accuracies of classification based on the proposed BioExpDNN in the selected three cases (i.e. CDS, CCBRDS, and HFCRDS) were 92.59%, 100%, and 78.9%, respectively. In the future, the proposed BioExpDNN can be adopted to extract features for other task types (e.g., regression, prediction, and so on) and other applications (e.g., social networks [35]).

ACKNOWLEDGMENT

This work was partially supported by the National Natural Science Foundation of China (Nos. 61906043, 61877010, 11501114, and 11901100), Fujian Natural Science Funds (No. 2019J01243), Funds of Education Department of Fujian Province (No. JAT190026) and Fuzhou University (Nos. 0330/50009113, 510872/GXRC-20016, 510930/XRC-20060, 510730/XRC-18075, 510809/GXRC-19037, 510649/XRC-18049, and 510650/XRC-18050).

Authorized licensed use limited to: Birkbeck University of London. Downloaded on March 28,2021 at 03:35:31 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] E. Pala, T. Denkceken, "Evaluation of miRNA expression profiles in schizophrenia using principal-component analysis-based unsupervised feature extraction method," *Journal of Computational Biology*, vol. 27, no. 8, pp. 1253-1263, 2020.

[2] M. Topolski, "The modified principal component analysis feature extraction method for the task of diagnosing chronic lymphocytic leukemia type B-CLL," *Journal of Universal Computer Science*, vol. 26, no. 6, pp. 734-746, 2020.

[3] L. Cheng, D. Li, G. Yu, Z. Zhang, X. Li, S. Yu, "A motor imagery EEG feature extraction method based on energy principal component analysis and deep belief networks," *IEEE Access*, vol. 8, pp. 21453-21472, 2020.

[4] X. Deng, X. Tian, S. Chen, C.J. Harris, "Deep principal component analysis based on layerwise feature extraction and its application to nonlinear process monitoring," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 6, pp. 2526-2540, 2019.

[5] X. Xiao, Z. Qiang, J. Zhao, Y. Qiang, P. Wang, P. Han, "A feature extraction method for lung nodules based on a multichannel principal component analysis network (PCANet)," *Multimedia Tools and Applications*, vol. 78, no. 13, pp. 17317-17335, 2019.

[6] A. Bellemans, G. Aversano, A. Coussement, A. Parente, "Feature extraction and reduced-order modelling of nitrogen plasma models using principal component analysis," *Computers & Chemical Engineering*, vol. 115, pp. 504-514, 2018.

[7] Y.H. Taguchi, "Tensor decomposition-based and principal-component-analysis-based unsupervised feature extraction applied to the gene expression and methylation profiles in the brains of social insects with multiple castes," *BMC Bioinformatics*, vol. 19, no. 4, p. 99, 2018.

[8] S.M.S. Shah, S. Batool, I. Khan, M.U. Ashraf, S.H. Abbas, S. A. Hussain, "Feature extraction through parallel Probabilistic Principal Component Analysis for heart disease diagnosis," *Physica A: Statistical Mechanics and its Applications*, vol. 482, pp. 796-807, 2017.

[9] Y. Ren, L. Liao, S.J. Maybank, Y. Zhang, X. Liu, "Hyperspectral Image Spectral-Spatial Feature Extraction via Tensor Principal Component Analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1431-1435, 2017.

[10] Y.H. Taguchi, "Identification of more feasible microRNA-mRNA interactions within multiple cancers using principal component analysis based unsupervised feature extraction," *International Journal of Molecular Sciences*, vol. 17, no. 5, p. 696, 2016.

[11] X. Song, M. Wang, H. Qiu, K. Li, C. Ang, "Auditory scene analysis-based feature extraction for indoor subarea localization using smartphones," *IEEE Sensors Journal*, vol. 19, no. 15, pp. 6309-6316, 2019.

[12] P. Zhu, C. Yin, Y. Cheng, X. Huang, J. Cao, C.M. Vong, P.K. Wong, "An improved feature extraction algorithm for automatic defect identification based on eddy current pulsed thermography," *Mechanical Systems and Signal Processing*, vol. 113, pp. 5-21, 2018.

[13] L. Yan, Y. He, L. Qin, C. Wu. D. Zhu. B. Ran, "A novel feature extraction model for traffic injury severity and its application to fatality analysis reporting system data analysis," *Science Progress*, vol. 103, no. 1, pp. 1-23, 2020.

[14] M. Iglesias Martínez, J. García-Gomez, C. Sáez, P. Fernández de Córdoba, J. Alberto Conejero, "Feature extraction and similarity of movement detection during sleep, based on higher order spectra and entropy of the actigraphy signal: results of the hispanic community health study/study of latinos," *Sensors*, vol. 18, no. 12, p. 4310, 2018.

[15] M.S. Akter, Md.R. Islam, Y. Iimura, H. Sugano, K. Fukumori, D. Wang, T. Tanaka, A. Cichocki, "Multiband entropy-based feature-extraction method for automatic identification of epileptic focus based on high-frequency components in interictal iEEG," *Scientific Reports*, vol. 10, no. 1, p. 7044, 2020.

[16] P. Huang, Z. Huang, X. Lu, Y. Cao, J. Yu, D. Hou, G. Zhang, "Study on glycoprotein terahertz time-domain spectroscopy based on composite multiscale entropy feature extraction method," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 229, p. 117948, 2020.

[17] F. Ding, C. Luo, "The entropy-based time domain feature extraction for online concept drift detection," *Entropy*, vol. 21, no. 12, p. 1187, 2019.

[18] C. Uyulan, T.T. Erguzel, N. Tarhan, "Entropy-based feature extraction technique in conjunction with wavelet packet transform for multi-mental task classification," *Biomedical Engineering / Biomedizinische Technik*, vol. 64, no. 5, pp. 529-542, 2019.

[19] N. Ji, L. Ma, H. Dong, X. Zhang, "EEG signals feature extraction based on DWT and EMD combined with approximate entropy," *Brain Sciences*, vol. 9, no. 8, p. 201, 2019.

[20] X. Chen, J. Chen, J. Liang, Y. Li, C.A. Courtney, Y. Yang, "Entropy-based surface electromyogram feature extraction for knee osteoarthritis classification," *IEEE Access*, vol. 7, pp. 164144-164151, 2019.

[21] J. Li, Y. Ying, C. Ji, "Study on Gas Turbine Gas-Path Fault Diagnosis Method Based on Quadratic Entropy Feature Extraction," *IEEE Access*, vol. 7, pp. 89118-89127, 2019.

[22] M. Li, R. Wang, J. Yang, L. Duan, An improved refined composite multivariate multiscale fuzzy entropy method for MI-EEG feature extraction," *Computational Intelligence and Neuroscience*, vol. 2019, p. 7529572, 2019.

[23] W. Fu, J. Tan, C. Li, Z. Zou, Q. Li, T. Chen, "A hybrid fault diagnosis approach for rotating machinery with the fusion of entropy-based feature extraction and SVM optimized by a chaos quantum sine cosine algorithm," *Entropy*, vol. 20, no. 9, p. 626, 2018.

[24] I. Mitiche, G. Morison, A. Nesbitt, B. Stewart, P. Boreham, "Entropy-based feature extraction for electromagnetic discharges classification in high-voltage power generation," *Entropy*, vol. 20, no. 8, p. 549, 2018.

[25] S. Nalband, A. Prince, A. Agrawal, "Entropy-based feature extraction and classification of vibroarthographic signal using complete ensemble empirical mode decomposition with adaptive noise," *IET Science, Measurement & Technology*, vol. 12, no. 3, pp. 350-359, 2018.

[26] S. Chen, Z. Luo, H. Gan, "An entropy fusion method for feature extraction of EEG," *Neural Computing & Applications*, vol. 29, no. 10, pp. 857-863, 2018.

[27] F. Khozeimeh, R. Alizadehsani, M. Roshanzamir, A. Khosravi, P. Layegh, S. Nahavandi, "An expert system for selecting wart treatment method," *Computers in Biology and Medicine*, vol. 81, pp. 167-175, 2017.

[28] F. Khozeimeh, F.J. Azad, Y.M. Oskouei, M. Jafari, S. Tehranian, R. Alizadehsani, P. Layegh, "Intralesional immunotherapy compared to cryotherapy in the treatment of warts," *International Journal of Dermatology*, vol. 56, no. 4, pp. 474-478, 2017.

[29] F. Khozeimeh, P. Layegh, R. Alizadehsani, M. Roshanzamir, " Cryotherapy data set," *UCI Machine Learning Repository*, 2018, Retrieved 15 Octo.ber 2020 from https://archive.ics.uci.edu/ml/datasets/Cryotherapy+Dataset+

[30] D. Sobar, R. Machmud, A. Wijaya, "Behavior determinant based cervical cancer early detection with machine learning algorithm," *Advanced Science Letters*, vol. 22, no. 10, pp. 3120-3123, 2016.

[31] D. Sobar, R. Machmud, A. Wijaya, " Cervical cancer behavior risk data set," *UCI Machine Learning Repository*, 2019, Retrieved 15 Octo.ber 2020 from https://archive.ics.uci.edu/ml/datasets/Cervical+Cancer+Behavior+Risk

[32] D. Chicco, G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Medical Informatics and Decision Making*, vol. 20, p. 16, 2020.

[33] D. Chicco, " Heart failure clinical records data set," *UCI Machine Learning Repository*, 2020, Retrieved 15 Octo.ber 2020 from https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records

[34] R. Lin, C.H. Chen, "An explainable feature extraction method based on deep learning techniques for social assimilation – a case study of cross-strait relations," *Basic & Clinical Pharmacology & Toxicology*, vol. 126(S4), pp. 274-275, 2020.

[35] H. Chen, H. Jin, S. Wu, "Minimizing inter-server communications by exploiting self-similarity in online social networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 4, pp. 1116-1130, 2016.

Authorized licensed use limited to: Birkbeck University of London. Downloaded on March 28,2021 at 03:35:31 UTC from IEEE Xplore. Restrictions apply.