DRNN: Deep Residual Neural Network for Heart Disease Prediction

Nianhao Xiao¹², Yuanchen Zou¹², Yaguang Yin³, Peishun Liu², Ruichun Tang³

¹ These authors contributed equally to this work.
² College of Information Science and Engineering, Ocean University of China, Qingdao, China
³ Academy of broadcasting science, Beijing, China
✉ tangruichun@ouc.edu.cn

Abstract. Heart disease is one of the major diseases threatening human health. This paper proposed a novel deep neural network model to predict heart disease based on routine clinical data. We adapt the deep residual structure to discover a novel Deep Residual Neural Network (DRNN). In order to verify the effectiveness of DRNN, we performed experiments on Heart Disease UCI. The accuracy reached 95%, which is better than the traditional machine learning methods among Random Forest 83%, Decision Tree 68%, Logistic Regression 87%, KNN 60%, Native Bayes 80%.

1. Introduction
Heart disease is one of the most harmful diseases owing to the highly lethality rate. Thus, heart disease prediction (HDP) is a significant and meaningful work. In the field of computer, HDP is regarded as a binary problem. According to a person's health data, we judge whether the person would suffer heart disease or not. That is to find the potential relationship between heart disease and human health data. Based on these studies[1][2], we can draw a conclusion that heart disease is mainly affected by gender, chest pain type etc.

The relationship between these factors and heart disease is not direct, but complex and nonlinear. A generous amount of data mining technologies[3][4] are applied to track this relationship. However, their performances are barely satisfactory. Deep Neural Network (DNN) shows excellent performance in finding complex nonlinear relations[5]. At the beginning, Deep Residual Network (ResNet)[6] is applied to image classification. ResNet can find the nonlinear relationship between images and labels by convolution of image. Meanwhile it can simplify the calculation of neural network through residual structure[7]. Inspired by this work, we propose DRNN. HDP method is based on the residual structure which can be used to diagnose heart disease. We get rid of the convolution layer of ResNet and directly replaced it with building the whole network with dense connection. And it is followed by a dropout layer[8], in order to prevent overfitting and gradient explosion caused by the connection.

This paper analysed and discussed the related work to HDP in Section II. Section III details the method and network structure. The results and analyses are given in Section IV. Section V summarizes the work.
2. Related Work

There are a great quantity of research using data mining on HDP. The existing methods of HDP can be roughly divided into two categories, traditional numerical methods and neural network methods.

(i) **Traditional Numerical Methods** are based on the basic principle of mathematics, most of which depend on the randomness of probability theory and statistical law. Although the calculation speed of this kind of method is fast, in current, it is difficult to improve the accuracy no matter how optimized the models. Dulhare et al[9] and Repaka et al[10] used naive Bayes to predict heart disease. Naive Bayes is based on the assumption that individuals are independent of each other. Raju et al[4] and Babu et al[1][11] compared the performance of KNN, SVM, genetic algorithm, random forest in HDP, among which SVM 87% is the best. Although they have analysed and optimized these methods, the bottleneck of the accuracy still exists. All the methods mentioned above are based on numerical calculation, and rely on random distribution. Besides there are also problems of low accuracy, researchers began to explore other methods.

(ii) **Neural Network Methods** are based on Bionics. A good heart surgeon can accurately judge whether a person has heart disease according to the clinical data. This method mimics the working mechanism of neurons in the human brain and trains neural networks to have the same judgment ability as cardiologists. Sonawane et al[11] presented a prediction system for heart disease using multilayer perceptron neural network (MLPNN). The model contained only one hidden layer with 100 neurons, but the accuracy was close to 90%. Kaiming et al[6] proposed Residual Networks (ResNet) which solved these problems. But ResNet cannot be applied to HDP directly for it was used to classify images. Neural network shows excellent performance in the field of data mining[12]. But limited by the gradient vanishing and gradient explosion of back propagation, the depth of the network model was constrained. We introduce the residual structure into the field of neural network, which makes the application of deeper and more complex neural network possible.

3. Method

3.1. Residual Structure

Unlike ResNet[6], we use dense connection layer replaces convolution layer. Then the Residual Structure (resblock) is constructed as follows. The shortcut connections simply perform identity mapping, and their outputs are added to the outputs of the stacked layers, as shown in the Fig.1(a).

Suppose that the final mapping of the network is $H(x)$. There is another mapping $F(x)$, satisfying $H(x) - x = F(x)$. Therefore, the original mapping $H(x)$ can be expressed as $F(x) + x$. The research shows that the structure has stronger feature extraction ability and simpler optimization ability.

3.2. Proposed Method based on Residual Structure

In this paper, we adopt resblock to build the DRNN for HDP. We propose a novel structure as shown in Fig.1(b). We defined the resblock as

$$y = \delta(F(x,\{W,b\}) + x):U$$

where $x$ and $y$ are the input and output of the resblock, $U$ is the number of units. The function $F(x,\{W,b\})$ indicate the mapping which the resblock to be learned. For the i-th resblock

$$F_i = \delta(W_i * x_{i}^{\text{input}} + b_i)$$

$$x_{i}^{\text{input}} = \delta(W_{i}^{\text{input}} * y_{i-1} + b_{i}^{\text{input}})$$

which $\delta$ denotes Relu , $W$ and $b$ is weight matrix and bias.
As shown in Fig.1(b), DRNN consists of an input layer followed by a regularization layer and a dropout layer. Then four resblocks are used to find the relationship mapping. Finally, the output layer outputs the prediction result. In summary, the loss function is

\[
J(W, b) = -\sum_i y_i' \log \phi(y_i'^p \{W, b, x\}) \\
+ (1 - y_i') \log(1 - y_i'^p \{W, b, x\})
\]  

(4)

Thus, the network can be expressed as

\[
\arg\min_{W, b} J(W, b) = -\frac{1}{N} \sum_i \sum_j y_j'^p \log \phi(y_j'^p \{W, b, x\}) \\
+ (1 - y_j') \log(1 - \phi(y_j'^p \{W, b, x\}))
\]  

(5)

Where \(N\) is the total number of samples for each batch, and \(C\) is the number of the categories. \(y_i'\) is the ground truth, \(y_i'^p\) is the probability of the category.

### 4. Experiment

#### 4.1. Environment and Settings

The hardware configuration of the experiment is, Intel Core (TM) i5-8400 CPU@2.8Ghz, 8G RAM, NVIDIA 1050ti graphics card. The operating system is Windows 10, using tensorflow 2.0 framework to build the whole network. We use the Adaptive Motion Estimation (adam) optimizer to minimize the loss function and set the initial learning rate to \(1 \times 10^{-3}\), \(\beta_1 = 0.9\), \(\beta_2 = 0.999\), \(\varepsilon = 1 \times 10^{-7}\), The number of iterations is 4000.

We apply the dataset named Heart Disease UCI[9], and encode the variables of string type with onehot (\(depth = 1\)), and remove the missing values. Through the correlation calculation, we get the correlation heatmap diagram (Fig.2(a)), and we can see that they all have a certain correlation. This is consistent with the previous research results[1][2][14]. We also analyze the Skewness of the dataset (Fig.2(b)), and the results are in line with the normal distribution, which shows that processing of the data rarely destroy the rationality of the dataset.
We use 80% of the dataset for model training and 20% for final testing. Experimental results show that the performance of DRNN is much higher than that of other machine learning methods (see Table 1). Specifically, the accuracy of DRNN is 95%, 12% higher than that of RF, 27% higher than DT, 18% higher than LR, 35% higher than KNN, 15% higher than NB, 27% higher than MLP.

5. Conclusion
This work applied deep residual structure to HDP task, and experiments show that our method is more accurate than other common machine learning methods. The effectiveness of deep residual structure in HDP is proved. We believe that this method is also applicable to other classification problems. The future research direction can be divided into two parts: (1) To explore the effect of different residual structures. For example, more resblocks or different number of nerve units. (2) To explore the application of depth residual structure in other classification problems.

Acknowledgments
This work is supported by the National Key Research and Development Program of China (2017YFC0806205). We are grateful to the anonymous reviewers for comments on the original manuscript.

| method         | class | precision | recall | f1-score | accuracy |
|----------------|-------|-----------|--------|----------|----------|
|                | 0     | 0.91      | 0.72   | 0.81     | 0.83     |
| Random Forest  | 1     | 0.78      | 0.94   | 0.85     |          |
| Decision Tree  | 0     | 0.75      | 0.52   | 0.61     | 0.68     |
|                | 1     | 0.65      | 0.84   | 0.73     |          |
| Logistic Regression | 0     | 0.59      | 0.55   | 0.57     | 0.87     |
|                | 1     | 0.61      | 0.65   | 0.62     |          |
| KNN            | 0     | 0.81      | 0.76   | 0.79     | 0.60     |
|                | 1     | 0.79      | 0.84   | 0.81     |          |
| Naive Bayes    | 0     | 0.81      | 0.76   | 0.79     | 0.80     |
|                | 1     | 0.79      | 0.84   | 0.81     |          |
| MLP            | 0     | 0.73      | 0.55   | 0.63     | 0.68     |
|                | 1     | 0.66      | 0.81   | 0.72     |          |
| Deep RES       | 0     | 0.91      | 1.00   | 0.95     | 0.95     |
|                | 1     | 0.97      | 0.92   | 0.96     |          |
References

[1] Babu, S., Vivek, E.M., Famina, K.P., Fida, K., Aswathi, P., Shanid, M., Hena, M.: Heart disease diagnosis using data mining technique. In: 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA). vol. 1, pp. 750–753 (April 2017). https://doi.org/10.1019/ICECA.2017.8203643

[2] Sowmiya, C., Sumitra, P.: Analytical study of heart disease diagnosis using classification techniques. In: 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS). pp. 1–5 (March 2017). https://doi.org/10.1109/ITCOSP.2017.8303115

[3] Jegan, C., Seenivasagam, V.: Review of heart disease prediction system using data mining and hybrid intelligent techniques. International Journal of Soft Computing 3, 2229–6956 (07 2013). https://doi.org/10.21917/ijsc.2013.0087

[4] Raju, C., Philipsy, E., Chacko, S., Padma Suresh, L., Deepa Rajan, S.: A survey on predicting heart disease using data mining techniques. In: 2018 Conference on Emerging Devices and Smart Systems (ICEDSS). pp. 253–255 (2018)

[5] Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, inception-resnet and the impact of residual connections on learning

[6] He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. arXiv e-prints arXiv:1512.03385 (Dec 2015)

[7] Jiao, J., Tu, W.C., He, S., Lau, R.W.H.: Formresnet: Formatted residual learning for image restoration. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2017)

[8] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 15, 1929–1958 (2014)

[9] Frank, A., & Asuncion, A. (2010). {UCI} Machine Learning Repository. Retrieved from http://archive.ics.uci.edu/ml Dulhare, U.: Prediction system for heart disease using naive bayes and particle swarm optimization. Biomedical Research 29 (01 2018)

[10] N., P., Kumar, P.R.: Usage of data mining techniques in predicting the heart diseases — naive bayes decision tree. In: 2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT). pp. 1–7 (April 2017). https://doi.org/10.1109/ICCPCT.2017.8074215

[11] Sonawane, J.S., Patil, D.R.: Prediction of heart disease using multilayer perceptron neural network. In: International Conference on Information Communication and Embedded Systems (ICICES2014). pp. 1–6 (Feb 2014). https://doi.org/10.1109/ICICES.2014.7033860

[12] Sethukkarasi, R., Kannan, A.: An intelligent system for mining temporal rules in clinical databases using fuzzy neural networks. European Journal of Scientific Research 70, 386–395 (02 2012)

[13] Schraudolph, N.: Centering neural network gradient factors 7700, 205–223 (01 2012). https://doi.org/10.1007/978-3-642-35289-8-14

[14] Palaniappan, S., Awang, R.: Intelligent heart disease prediction system using data mining techniques. In: 2008 IEEE/ACS International Conference on Computer Systems and Applications. pp. 108–115 (March 2008). https://doi.org/10.1109/AICCSA.2008.4493524