On Machine Learning Models for Heart Disease Diagnosis

Chu-Hsing Lin*1, Po-Kai Yang2, Yu-Chiao Lin3, Pin-Kuei Fu4

1,2Department of Computer Science, Tunghai University, Taichung, Taiwan
3Department of Chinese Medicine, China Medical University, Taichung, Taiwan
4Department of Critical Care, Taichung Veterans General Hospital, Taiwan
{*1 chlin; 2 pky0023}@thu.edu.tw, 3 u107070301@cmu.edu.tw, 4 yetquen@gmail.com

Abstract

Convolutional Neural Networks (CNNs) have different architecture than regular Neural Networks (NNs) and are both applied extensively in many application fields. In this article, we used both of the two machine learning models in the heart disease diagnosis problems. We implemented the algorithms, tuned the parameters, and conducted a series of experiments. We aim to compare the prediction accuracy of the two models under different parameters settings. We used the Cleveland database which is took from UCI learning dataset repository for diagnosis heart disease. From the experimental results, we found that NNs outperform CNNs in prediction accuracy in most of the cases.

Key words: Heart Disease Diagnosis, Convolutional Neural Networks, Neural Networks, Convolution Layer, Pooling Layer, Prediction Accuracy.

Introduction

According to the investigation made by the World Health Organization in 2016, heart diseases play the most important role in top 10 global causes of deaths, to which ischemic heart disease (IHD) contributed most while stroke is the second one. Thus, how to diagnose heart disease fast and precisely is a pivotal issue. Lots of studies have been made to enhance the accuracy of finding heart diseases. In the previous work, the accuracy was almost between 61% [2] and 93% [3] approximately while the best results are 98% accuracy in MLP model and 96% accuracy in SVM model [1].

In this article, we used the Cleveland Heart Disease Data Set from the UC Irvine Machine Learning Repository as the training and test data. [4] The Cleveland dataset is very popularly used in academy for heart disease diagnosis studies. The dataset contains 303 instances and only 14 attributes are used from UCI dataset repository. Table I shows the 14 attributes description of the dataset. As shown in Table I, the fourteenth attribute is the predicted attribute that indicates whether the heart is normal or not. Two fundamental outputs are recognized where the value “0” denotes a healthy heart and “1” for sick one.

Convolutional Neural Networks (CNNs) have different architecture than regular Neural Networks (NNs) and are both applied extensively in many application fields. In this article, we used both of the two machine learning models in the heart disease diagnosis problems. Based on the Cleveland Heart Disease Data Set, we aim to compare the performance of the two machine learning models.

TABLE I
ATTRIBUTE DESCRIPTION

Age	Numerical
Sex	1: male; 0: female
CP	Chest pain type -- 1:typical angina; 2: atypical angina; 3: non-angina pain; 4: asymptomatic
Trestbps	Resting blood pressure in mm Hg
Chol	Serum cholesterol in mg/dl
Fbs	Indicator if fasting blood sugar > 120 mg/dl -- 1: true ;0: false
Restecg	Resting electrocardiographic hic results -- 0: normal; 1: having ST-T wave abnormality; 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
Thalach	Maximum heart rate achieved
Exang	Indicator of whether the angina is exercise induced -- 1: yes; 0: no
Oldpeak	ST depression induced by exercise relative to rest
Slope	The slope of the peak exercise ST segment -- 1: upsloping; 2: flat; 3: downs loping
Ca	Number of major vessels (0-3) colored by fluoroscopy
Thal	Summary of heart condition -- 3: normal; 6: fixed defect; 7: reversible defect
Num	The predicted attribute -- 0: healthy; 1: sick

We designed and implemented the programs for both models, tuning the parameters, and conducted a series of experiments. We calculated the prediction accuracy of the two methods under different parameters settings. From the experimental results, we found that, for the CNN model, we obtained stably about 80% of accuracy by adjusting the number of neurons in each layer and the number of layers. While, for the NN model, we obtained 93 % of accuracy in the case of two hidden layers and the number of neurons were less than 20 in each layer.

The organization of this paper is as follow. In Section 2, we describe the background knowledge of the paper. The design methodology and experiment setting appear in Section 3. In Section 4, we present the experiments and the results. Finally, we give a conclusion.

Background

A. Neural Networks

Regular Neural Networks (NNs) consists of an input layer, several hidden layers, and an output layer. [5-6] It transforms an input array by putting it into the input layer, going through a series of hidden layers, and finally obtains the prediction result

ISBN: 978-1-7281-8712-9

Authorized licensed use limited to: University of Zagreb. Downloaded on April 27,2025 at 11:04:10 UTC from IEEE Xplore. Restrictions apply.

from the output layer. Every layer in an NN is made up of a set of neurons, where each layer is fully connected to all neurons in the previous layer. Suppose a neuron with n inputs, every input X_i is multiplied by the respective weight W_i, computing the summation and go through the activation function, then the result is output.

B. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have a different architecture comparing with regular NNs in several aspects. [7-8] First, a CNN arranges its neurons in each layer with three dimensions (width, height, depth). Second, the neurons in one layer do not connect to all the neurons in the previous layer. In addition, CNNs have two component parts: the feature extractor and the classifier. In the feature extractor part, the network will perform a series of convolutional (by convolution layers) and pooling (by pooling layers) operations by which the features are detected. In the classifier part, the fully connected layers (FCNet) serve as a classifier on the extracted features. And it then also assigns a probability to the input array to indicate what it predicts.

Methodology

We designed algorithms for both of the two machine learning models, NN and CNN, in the heart disease diagnosis problems. Based on the Cleveland Heart Disease Data Set, we used 80% of the dataset as the training data and the rest 20% as the test data. We implemented programs for both models, tuning the parameters, and conducted a series of experiments. We calculated the prediction accuracy of the two methods under different parameters settings. The details of the experiments and the results will appear in the next section.

We used a statistical measure to compute the accuracy of a classifier. Accuracy is the ratio of the number of correct predictions to the number of total tests. That is, it is the proportion of true results, including the true positive and the true negative, in the tests. Accuracy can be calculated as the following equation:

$$Accuracy = \frac{T_P+T_N}{T_P+T_N+F_P+F_N} \quad (1)$$

where

T_P (True Positive): the number of test cases that have heart disease and proper diagnoses. T_N (True Negative): the number of test cases that are healthy and properly diagnosed. F_P (False Positive): the number of test cases that are healthy but wrongly sick diagnosed. F_N (False Negative): the number of test cases that have heart disease but wrong diagnoses.

The data used in this paper comes from UCI Machine Learning Repository [4], which contains 303 instances and only 14 attributes are used.

For the experimental environment, the Operating System is 64-bit Ubuntu 16.04 LTS, the processor is: Intel® Core™2 Quad CPU Q8200 @ 2.33GHz × 4, and the memory is 4G. The programming languages we used are: Python 3 and TensorFlow.

Experiments and Results

A. Experiment 1: Categorical Transformation

Categorical transformation is to transfer the categorical features into numerical features. [9] For example, the Sex feature in categorical type, "Male/(1)" and "Female/(0)," is transferred to "(1,0)" and "(0,1)", in numerical type, respectively. In this experiment, we concern whether the categorical transformation influence the accuracy of heart disease prediction or not.

We first used the NN model and variated the numbers of hidden layers, with the neuron number greater than 100 in each layer. Table II and Fig. 1 show the accuracy comparison for NN, concerning whether categorical transformation is included or not.

TABLE II
ACCURACY OF NN ON CATEGORICAL

Number of Hidden Layers	With Categorical.	Without Categorical.
1	76.70%	74.75%
2	75.72%	79.61%
3	70.87%	73.78%

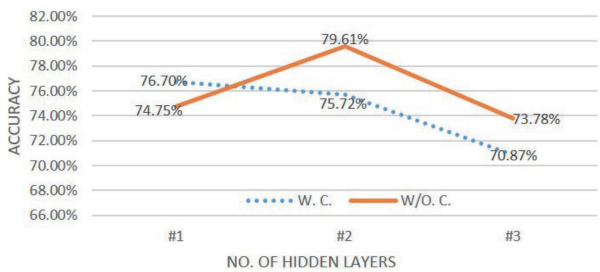

Fig. 1. Accuracy of NN on Categorical

Next, we used 2-layers of convolution and pooling in the CNN model and variated the numbers of hidden layers in the fully connected network (FCNet), with the neuron number greater than 100 in each layer. Table III and Fig. 2 show the accuracy comparison for CNN, concerning whether categorical transformation is included or not.

TABLE III
ACCURACY OF CNN ON CATEGORICAL

Number of Hidden Layers in the FCNet	With Categorical.	Without Categorical.
1	80.60%	80.60%
2	79.61%	81.55%
3	77.67%	76.69%

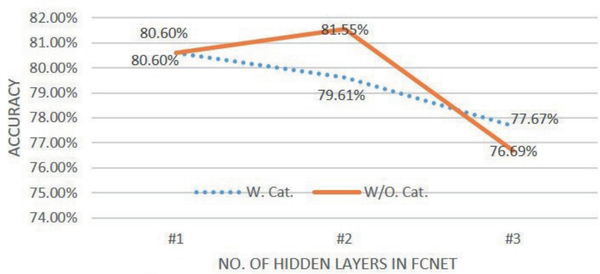

Fig. 2. Accuracy of CNN on Categorical

Authorized licensed use limited to: University of Zagreb. Downloaded on April 17,2025 at 11:04:10 UTC from IEEE Xplore. Restrictions apply.
ISBN: 978-1-7281-8742-9

As in Table Ⅱ and Ⅲ, we can obtain the follows:

(1) From Table Ⅱ, we can see that the accuracy is higher without categorical transformation in NN. It is probably because the sparsity of non-zero values in the numerical data after transforming, the features become less important.

(2) From Table Ⅲ, transforming does not affect the results of CNN much. It is probably because the transformed features are preserved after convolution layers and pooling layers.

(3) In both NN and CNN, the accuracy are the best in the cases of two hidden layers and the worst when three hidden layers are applied.

B. Experiment 2: Imbalanced data (classes) in training dataset

To increase the accuracy of machine learning, we balanced the classes of the training data. [9-10] That is, we use the same numbers of data from the two classes as training data so that the learning model does not bias towards a certain result. In *Experiment 1*, the numbers of the two classes are 113 with heart disease and 87 without heart disease (imbalanced data approximately 11 to 9). *In Experiment 2*, the data is balanced, the numbers of the two classes for the training data are both 100.

Table Ⅳ and Fig. 3 show the comparing results after classes balancing of training data using the same NN model as in *Experiment 1*.

TABLE IV
ACCURACY OF NN ON BALANCING

Number of Hidden Layers	With Balancing	Without Balancing
1	81.44%	74.75%
2	85.50%	79.61%
3	86.40%	73.78%

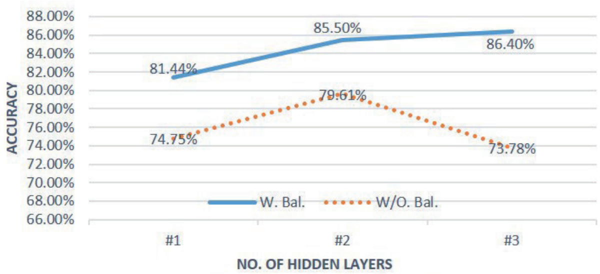

Fig. 3. Accuracy of NN on Balancing

Table V and Fig. 4 show the comparing results after classes balancing of training data using the same CNN model as in *Experiment 1*.

TABLE V
ACCURACY OF CNN ON BALANCING

Number of Hidden Layers	With Balancing	Without Balancing
1	80.60%	80.60%
2	82.52%	81.55%
3	78.65%	76.69%

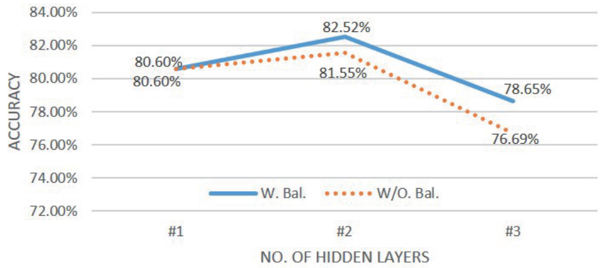

Fig. 4. Accuracy of CNN on Balancing

As in Table Ⅳ and Ⅴ, we can have the following result.

(1) In Table IV, in all cases the accuracy of the NN model has been improved. It is assumed that in the case of a small amount of data, class balancing has a greater impact on the NN learning model.

(2) In Table Ⅴ, though the accuracy of the CNN model has been improved, the effect is not as significant as the NN model.

C. Experiment 3: Variant number of layers and neurons

The purpose of this experiment is to test the correlation among the number of neurons, the number of hidden layers and the accuracy of the learning model. In Table VI, it showed the comparison of NN model among the best results of variant number of hidden layers, from 1 to 3, with more than 100, within 21 to 100, and 20 or fewer neurons per layer. Fig. 5 depicts the comparison and the trend of accuracy for data showed in Table VI.

TABLE VI
ACCURACY OF NN ON VARIANT NUMBERS OF HIDDEN LAYERS AND NEURONS IN EACH LAYER

Number of Neurons	1 Hidden Layers	2 Hidden Layers	3 Hidden Layer
> 100	81.44%	85.5%	86.40%
21~100	87.37%	88.35%	83.50%
<= 20	91.26%	93.81%	82.52%

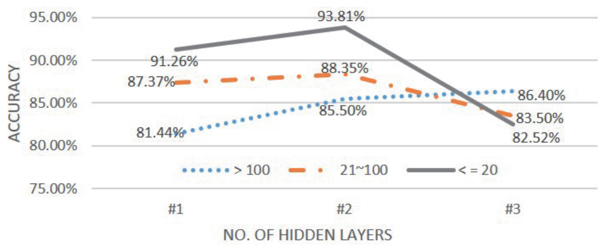

Fig. 5. NN on variant numbers of hidden layers and neurons in each layer

Similarly, Table VII and Fig. 6 show the best results for CNN model of variant number of hidden layers in the FCNet, from 1 to 3, with more than 100, within 21 to 100, and 20 or fewer neurons per layer.

ISBN: 978-1-7281-8712-9
Authorized licensed use limited to: University of Zagreb. Downloaded on April 17,2025 at 11:04:10 UTC from IEEE Xplore. Restrictions apply.

TABLE VII
ACCURACY OF CNN ON VARIANT NUMBERS OF HIDDEN LAYERS IN FCNET AND NEURONS IN EACH LAYER

Number of Neurons	1 Hidden Layers	2 Hidden Layers	3 Hidden Layer
> 100	80.60%	82.52%	78.65%
21~100	80.60%	80.60%	76.70%
< = 20	83.50%	82.52%	79.30%

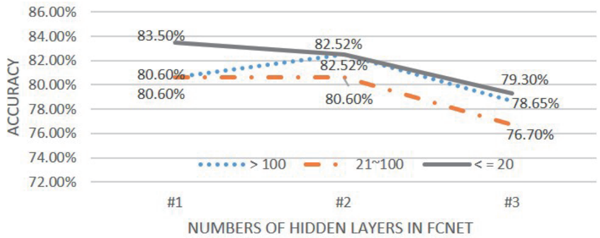

Fig. 6. CNN on variant numbers of hidden layers in FCNet and neurons in each layer

We can have the follows from Table VI and Table VII:

(1) By fixing the number of neurons per layer and changing the number of hidden layers, it can be found that in the cases of two hidden layers both of the two models have better performances.

(2) We got better result when the number of neurons lesser than 20, by fixing the number of hidden layers and changing the number of neurons in each layer.

(3) From the results, it seemed no absolute conclusion of the accuracy rate and the number of hidden layers and neurons. There are several factors to influence the optimal number of neurons. We found that 20 or fewer neurons will be better in this cases because of the overfitting problem when larger number of neurons.

Conclusions

In this article, we used both of NN and CNN machine learning models to solve the heart disease diagnosis problems. Using the Cleveland Heart Disease Data Set, we compare the performance of the two machine learning models by adjusting the parameter settings and conducted a series of experiments. Summarizing the experimental results above, the following conclusions and discussions were made:

(1) Though the number of hidden layers and the number of neurons per layer are crucial to machine learning models, there is no general rule to determine the numbers for the best performance. Usually, it is heuristic and may also be affected by many factors such as the number of training data and input features.

(2) In the case of not large enough of training data, the imbalanced-classes problem can result in a bias outcome for NN model. We need to make the training data balanced in the classes.

(3) The accuracy of CNN model is not as high as that of NN though, in almost all the cases it behaved stably in performance. For that CNN underperforms NN, one possible reason we assumed is the size of dataset only 303 instances. For this, we shall investigate the models based on larger dataset in the near future.

Acknowledgement

This research was partly supported by the Ministry of Science and Technology, Taiwan, under grant number MOST 107 - 2221 - E - 029 - 005 - MY3.

References

[1] Tabreer T. Hasan, Manal H. Jasim and Ivan A. Hashim, “Heart Disease Diagnosis System based on Multi-Layer Perceptron neural network and Support Vector Machine,” International Journal of Current Engineering and Technology, 2017.

[2] S. Bhatia, P. Prakash and G. N. Pillai, “SVM Based Decision Support System for Heart Disease Classification with Integer-Coded Genetic Algorithm to Select Critical Features,” in Proceedings of the World Congress on Engineering and Computer Science, San Francisco, USA, 2008.

[3] X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang and Q. Wang, “A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method,” Computational and Mathematical Methods in Medicine, pp. 11, 2017.

[4] UCI Machine Learning Repository website: http://archive.ics.uci.edu/ml/datasets/Heart+Disease

[5] J. A. Anderson, An Introduction to Neural Networks, Prentice Hall, 2003.

[6] P. Louridas, C. Ebert, Machine Learning, IEEE Computer Society, pp.110-115, 2016.

[7] Keiron O’Shea1 and Ryan Nash, An Introduction to Convolutional Neural Networks, arXiv:1511.08458v2, Dec. 2015.

[8] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, “Deep Learning,” Nature, 521: pp.436–444, 2015.

[9] Henrik Brink, Joseph W. Richards, Mark Fetherolf, Real-World Machine Learning, Manning, 2017.

[10] V. López, A. Fernandez, S. Garcia, V. Palade and F. Herrera, “An Insight into Classification with Imbalanced Data: Empirical Results and Current Trends on Using Data Intrinsic Characteristics,” Information Sciences, 250, pp.113-141, 2013. doi: 10.1016/j.ins.2013.07.007

Authorized licensed use limited to: University of Zagreb. Downloaded on April 17,2025 at 11:04:10 UTC from IEEE Xplore. Restrictions apply.