Critical Input Data Channels Selection for Progressive Work Exercise Test by Neural Network Sensitivity Analysis

Avni H. Rambhia[†] , *Robb Glenny* [‡] , *Jenq-Neng Hwang* [†]

Box # 352500, Dept. of Electrial Engineering[†]
Box # 356522, Dept. of Medicine[‡]
University of Washington
Seattle, WA 98195

Abstract

We aimed at training a neural network to classify stress test exercise data into one of three classes: normal, heart failure, or lung failure. Good classification accuracy was obtained using a backpropagation neural network architecture with one hidden layer during cross validation on data set of 110 vectors, when all 17 channels were used. We further aimed at determining which of these channels were critical to the decision making process. This was done through an input sensitivity analysis. Results showed that nine channels formed a critical superset of which possibly any eight could achieve almost perfect classification. We thus show that faster and more accurate classification may be obtained by input channel elimination due to dimension reduction of input space, which makes better generalization.

1. INTRODUCTION TO EXERCISE DATA

Neural Networks (NNs) have been successfully used to classify and recognize patterns of response in medical diagnoses, such as the Progressive Work Exercise Test (PWET) [1,2] where a patient is asked to exercise on a stationary bicycle while an array of electrodes and sensors recording vital health information. The exercise data were obtained from three different subject classes: *normal subjects*, *subjects with cardiac failure*, and *subjects with Interstitial Lung Disease*. The pattern of each subject's response to exercise was characterized by a third degree polynomial to each of 17 different data channels as a function of wattage. The 68 coefficients from each subject were used as inputs to a multilayer perception NN classifier.

The 17 input data channels are referenced by their respective numbers throughout this paper [1].

1. **Watts:** Power generated by the subject on the bike. All other channels are interpreted (regressed) based on the watts being spent by the patient.
2. **Oxygen Consumption per minute (VO_2):** This is used to quantify the degree of impairment of the subject. VO_2 bears a linear relationship to watts, hence the two may be considered surrogates of each other.
3. **VO_2 /kg:** Channel 2 scaled by body weight.
4. **Respiratory Quotient (R):** Ratio of CO_2 production to O_2 consumption. At rest, it is about 0.8 and rises with exercise. The point where it crosses 1 is critical.
5. **Ventilatory Equivalent for Oxygen consumption (VEO2):** Computed as VO2/VE.
6. **Ventilatory Equivalent for Carbon Dioxide consumption (VECO2):** Computed as VCO2/VE.
7. **Minute Ventilation (VE):** The amount of air breathed per minute.
8. **VCO_2**: the amount of carbon dioxide produced per minute.
9. **Vd/Vt:** Vd denotes dead space. Vt denotes tidal volume.
10. **The tidal volume (Vt):** The volume of each breath.

11. **Ti/Ttot:** This is the ratio of time taken to breathe in to the total breath cycle time.
12. **Respiratory Rate (RR):** This shows the same behaviour as channel 10.
13. **End tidal CO2 ($EtCO_2$):** This is the concentration of CO_2 exhaled in each breath.
14. **Oxygen Saturation (SO_2):** This is the percent of hemoglobin that carries oxygen.
15. **(VO_2/HR):** This is the amount of oxygen consumed per heartbeat and is calculated as its name suggests.
16. **Blood Pressure.** This is sensitive to heart disease.
17. **Heart Rate (HR):** Diseased hearts, depending on type of disease, may show one of two abnormalities:
    - HR starts high and then gets closer to the theoretical maximum over time.
    - Chronotropic incompetence is shown, meaning that the heart rate cannot rise high enough.

2. DATA CHANNEL ELIMINATION

Having trained a reasonably good neural network classifier [3] using the 17 data channels data as the inputs, an input ***sensitivity analysis*** [4] is conducted on the trained network, using the training data. Sensitivity of each of the three outputs to each of the 68 inputs (17 channels with 4 coefficients of the 3rd order polynomial fit to each) is calculated as a partial derivative of the output with respect to the input [4] (assuming a 2-layer, one hidden layer network).

$$\frac{\partial y_i}{\partial x_m} = y_i[1-y_i]\sum_{j=1}^{Nh} w_{ij}(2)w_{jm}(1)a_j(1)[1-a_j(1)]$$

where

x_i denotes the i-th input to the network.

a_i (l) denotes the i-th element of the hidden layer of the network. Since it is a two-layer network, $y_i = a_i$ (2).

N_h = number of neurons in the hidden layer.

N_i = number of neurons in the input (first) layer.

w_{jm}(l) denotes the weight of the connection from the i^{th} node of the (l-1) layer to the j^{th} node of the l^{th} layer.

Once the average sensitivities (over all data in the same class) are known, channels with a low average sensitivity across all classes can be eliminated as unimportant to the decision making process. The step-by-step process is given below.

1. Train a neural network on all available data.
2. Using these weights, compute the sensitivity for each vector.
3. Group vectors according to their classes.
4. Compute the mean and absolute sensitivities for each class.
5. Determine the channel that has least sensitivity across classes.
6. Eliminate this channel and cross-validate.
7. If there is no degradation of response, the channel is indeed superfluous.
8. Steps 1-7 are repeated successively for smaller size of data channels, till a point of classification degradation is reached. Now, no more channels can be removed..

3. SIMULATION RESULTS

There are 111 data vectors available, which were broken up into 11 files of 10 vectors each. Thus, we have eleven tests during cross validation, each testing with 100 training vectors and 10 testing vectors. All the results were obtained based on an one-hidden-layer network with 20 hidden neurons. Reported below are the classification accuracy (in terms of %) for each disease class (N,C,P), taking into account all 11 cross-validation sets.

Starting with all 17 data channels (68 inputs), we gradually eliminate one input data channel (4 inputs) at each trial with least sensitivity as shown in Table 1. A breakdown of the classification accuracy can be clearly observed when two channels were eliminated after Trial 9. More specifically, out of 17 data channels, 9 of them (channels 4,7,8,9,10,11,12,14,15) were involved in the decision making process, while at least 8 of them are critically required for better performance (from 95.4% using all 17 channels to 98.2% using 9 channels). This better performance was achieved due to the lower dimensional input space ensuring a better generalization capability in the classification.

Table 1: The sequence of data channel Elimination (one at a time).

Trial	Additional channel deleted
1	Base case
2	1
3	6
4	5
5	16
6	13
7	17
8	2
9	3
10	4
11	7

Trial 1 (all 17 channels):

class	1	2	3	4	5	6	7	8	9	10	11
N	100	100	100	100	100	100	100	100	100	100	100
C	75	100	100	100	100	100	75	100	100	100	100
P	100	66.7	100	100	80	100	75	100	100	50	100

Trial 2 & 3 (after 1st and 6th channels are removed):

class	1	2	3	4	5	6	7	8	9	10	11
N	100	100	100	100	100	100	100	100	100	100	100
C	75	100	100	100	100	100	75	100	100	100	100
P	100	66.7	100	**83.3**	80	100	75	100	100	50	100

Trial 4 & 5 (after 5th and 16th channels are further removed):

class	1	2	3	4	5	6	7	8	9	10	11
N	100	100	100	100	100	100	100	100	100	100	100
C	**100**	100	100	100	100	100	75	**75**	100	100	100
P	100	66.7	100	**100**	80	100	75	100	100	50	100

Trial 6 & 7 (after 13th and 17th channels are further removed):

:

class	1	2	3	4	5	6	7	8	9	10	11
N	100	100	100	100	100	100	100	100	100	100	100
C	100	100	100	100	100	100	75	75	100	100	100
P	100	**100**	100	100	80	**50**	75	100	100	100	100

Trial 8 (after 2nd channel is removed):

class	1	2	3	4	5	6	7	8	9	10	11
N	100	100	100	100	100	100	100	100	100	100	100
C	100	100	100	100	100	100	75	**100**	100	100	100
P	100	100	100	100	80	50	75	100	100	100	100

Trial 9 & 10 (after 3rd and 4th channels are removed):

class	1	2	3	4	5	6	7	8	9	10	11
N	100	100	100	100	100	100	100	100	100	100	100
C	100	100	100	100	100	100	**100**	100	100	100	100
P	100	100	100	100	**100**	**100**	**50**	100	100	100	100

Trial 11 (after 7th channel is removed):

class	1	2	3	4	5	6	7	8	9	10	11
N	100	100	100	100	**33**	100	100	100	100	100	100
C	100	100	100	100	100	100	**75**	100	100	100	**83.3**
P	100	100	100	100	100	**50**	50	100	100	100	100

Note also that of the channels that were deemed redundant to the decision process, most contained values that were deducible from retained channels (VECO2, VEO2, VCO2/VO2, etc.). Some channels contained values that differed within a given class (HR), and some were obtained by manual measurement at predetermined intervals, followed by linear interpolation (HR, BP). In both the latter cases, it is possible that inconsistencies within a class caused a reduction in sensitivity.

4. CONCLUSION

Neural networks were useful to automatically diagnose conditions of normal, heart failure and lung failure from exercise data. Sensitivity analysis was found useful to isolate critical channels, thereby reducing the dimension of the decision space and increasing speed and accuracy of the classification system.

Acknowledgement: The authors would like to show their deep appreciation to Mr. Tri Vo of Adobe Inc. who helped in setting up the simulations for this paper.

REFERENCES:

1. Kian Shahla, "Neural Networks Applied to Exercise Data," EE499 Technical Report, University of Washington, August 1995.
2. Robb W. Glenny, Jenq-Neng Hwang, and H. T. Robertson, "Interpretation of Progressive Work Exercise Tests (PWET) using a Neural Network for Pattern Recognition," Am. J. Respir. Crit. Care Med. 153:A20, May 1996.
3. D. E. Rumelhart, G. E. Hinton, R. J. Williams, Parallel Distributed Processing, Vol. 1, Chap. 8, pp. 318-362, MIT Press, Cambridge, MA, 1986.
4. J. N. Hwang, J. J. Choi, S. Oh, R. J. Marks, "Query Based Learning Applied to Partially Trained Multilayer Perceptron," IEEE T-NN, 2(1):131-136, Jan. 1991.