Fuzzy Preprocessing Rules for the Improvement of an Artificial Neural Network Well Log Interpretation Model

Kok Wai Wong
School of Information Technology
Murdoch University
South St, Murdoch
Western Australia 6150
email: kwong@murdoch.edu.au

Chun Che Fung and Kok Way Law
School of Electrical and Computer Engineering
Curtin University of Technology
Kent St, Bentley
Western Australia 6102
email: tfungcc@curtin.edu.au

Abstract: The success of an Artificial Neural Network (ANN) based data interpretation model depends heavily on the availability and the characteristics of the training data. In the process of developing a reliable well log interpretation model, a log analyst has to spend many hours to perform pre-processing on the training data set. This demands substantial experience and expertise from the analyst. This paper proposes a fuzzy logic approach to integrate the knowledge of the log analysts in the stage of pre-processing. This paper also presents results from an experimental study which demonstrated the implementation of the fuzzy preprocessing technique which has increased the prediction accuracy of the ANN well log interpretation model. This new method has the potential to be a useful and important tool for the professional well log analysts.

Keywords
Neural Networks, Well log data, fuzzy rules, data analysis

I. INTRODUCTION

Well logging plays an essential role in the determination of the production potential of a hydrocarbon reservoir [1]. It is a geophysical prospecting technique that has been in use since 1927. The process involves lowering a number of instruments into a borehole with the purpose of collecting data at different depth intervals. The measurements broadly fall into three categories: electrical, nuclear and acoustic. A log analyst is one who interprets the data with an objective to translate the log data into petrophysical parameters of the well. To obtain an accurate picture of the important petrophysical parameters, extensive analysis of the core has to be carried out. This will provide answers to the questions on the petrophysical properties of the particular borehole such as lithology, porosity, amount of clay, grain size, water saturation, permeability and many others. All these answers are essential to the evaluation of the reservoir formation [2].

Although core data obtained from the detailed laboratory analysis are deemed to be most accurate, the analysis process is an expensive and lengthy exercise. Usually, limited core data are available at certain intervals. They are used as the basis to establish an interpretation model for other zones with similar log responses. Ideally, the model could be used to interpret log data from wells within the neighbouring region without the need to carry out further core analysis. This requires an integrated knowledge of the tool responses and understanding of the geology of the region, together with various mathematical techniques in order to derive an interpretation model which relates the log data to the petrophysical properties. However, the establishment of an accurate well log interpretation model is not an easy task due to the complexity of different factors that influence the log responses.

Applications of Artificial Neural Network (ANN) to well log interpretation have been reported and have shown to be successful in recent years. In particular, the Backpropagation Neural Networks (BPNN) are the most widely used [3,4,5,6,7]. Normally, well log data such as gamma ray (GR), bulk density (RHOB), sonic travel time (DT) and Resistivity (RT) are used as the inputs to the network. Corresponding core data at the same depth of specific petrophysical properties, such as permeability and porosity are used as the outputs.

Apparently, the BPNN-based method is a promising solution to the problem. However, the BPNN model becomes unreliable if the preprocessing of the training data is not handled properly. The training data plays a very important role in the success of the BPNN well log interpretation model. It is the set of training data that determines the underlying function, which is supposed to be learned by the BPNN model. Usually, log analysts have to spend many hours to perform tedious pre-processing tasks on the training data set before it can be used. The two main aims of this paper are (1) to make use of Computer Intelligence (CI) techniques to simplify the manual process; and (2) to incorporate the experience or knowledge of the log analyst in a format that could be used to improve the prediction accuracy of the analysis.

In this paper, the derivation of a Fuzzy Pre-processing technique is proposed to improve the performance of a BPNN well log interpretation model. With the past experience on petrophysics theory and knowledge of the wells, log analysts normally have some heuristic rules and expectations from the well log interpretation results. For example, a log analyst knows that the permeability should be high if the response of the gamma ray is low. However, in most BPNN training, this knowledge of the log analyst is not incorporated in the learning process. In the worst case, if there are errors in the training data, the BPNN may not be able to realise a reasonable prediction model. Since fuzzy rules are expressed in linguistics terms close to human knowledge, it will be easier for the log analyst to incorporate their knowledge in a fuzzy rule formats. The fuzzy rules created can then be used to assist the interpretation process during the pre-processing stage.

II. IMPORTANCE OF PREPROCESSING

Of all the different types of ANN configuration, BPNN is by far the most popularly used in building the well log interpretation model. This is mainly due to its ability of performing good function approximation or generalization characteristic. In function approximation, the BPNN is similar and comparable to non-parametric estimators in statistics [8, 9]. The objective is to build a model to represent the relationship between the input data set x and the target data set y without any assumed prior parameters. Given that the input vector X and the target vector Y, expression (1) can be used to describe the relationship:

$$Y = g(X) \qquad (1)$$

When obtaining the training set, there will be some environmental factors that affect the measurements. Therefore it is not possible to have an exact function, $g(\ )$, that describes the relationship between X and Y. However, a probabilistic relationship governed by a joint probability law $P(v)$ can be used to describe the relative frequency of occurrence of vector pair (X_n, Y_n) for n training data. The joint probability law $P(v)$ can be further separated into an environmental probability law $P(\mu)$ and a conditional probability law $P(\gamma)$. For notation expression, the probability law is expressed as:

$$P(v) = P(\mu)P(\gamma) \qquad (2)$$

The environmental probability law $P(\mu)$ describes the occurrence of the input X. The conditional probability law $P(\gamma)$ describes the occurrence of the output Y based on the given input X. A vector pair (X, Y) is considered as noise if X does not follow the environmental probability law $P(\mu)$, or the output Y based on the given X does not follow the conditional probability law $P(\gamma)$.

From (1), the relationship $g(X)$ based on the available training set can be assumed to be analogous to the conditional probability law $P(\gamma)$. Therefore, the BPNN is performing the role of estimating $P(\gamma)$. It can also be denoted as $E(Y|X)$ as the Expectation of Y given X.

Therefore:

$$g(X) = E(Y \mid X) \qquad (3)$$

In BPNN, $g(X)$ is not always obtained directly from the training set (X_n, Y_n). It has to undergo certain training process in realising the best $g(X)$. In a BPNN, the best $g(X)$ model is directly related to the internal weights W, which can be expressed as:

$$g(X) \approx f(X, W^*) \qquad (4)$$

where W^* denotes the set of the weights giving the best prediction;
and $f(\ )$ is the estimating function of the network.

From the above condition and taking the error into account, equation (1) is therefore:

$$Y = f(X, W^*) + \theta \qquad (5)$$

where θ denotes the error.

The output vector (predicted value), O will then be:

$$O = f(X, W) \qquad (6)$$

To find the best weights W^* so as to minimise the error function θ, a BPNN makes use of the error backpropagation learning algorithm [10] to perform the mean square errors minimisation process,

$$\sum_{i=1}^{n}[Y - f(X,W)]^2 \text{, or } \sum_{i=1}^{n}[Y - O]^2 \qquad (7)$$

With the above analysis, it is observed that the BPNN can generalise from noisy data by estimating $P(\gamma)$. It can then be argued that preprocessing of the training data may not be necessary. However, as a BPNN makes use of the backpropagation learning algorithm to estimate the function, any corrupted or irrelevant data will widen the search space in the data analysis. In the worst case when the distribution of the training data may not be able to distinguish noisy data, the final interpretation model will also provide unreasonable prediction. At this point, it can be deduced that noisy or corrupted data has two effects: (1) it may slow down the training time, and (2) it may distort the actual estimation function.

Preprocessing is therefore necessary, as any noisy data may also weaken the predictive capability of the BPNN. Preprocessing will allow log analysts to decide on how to present the training data to the BPNN and to lead the network to learn what is desired. As the log analysts know what should be the most desirable features for the final interpretation model, preprocessing will therefore enhance the data analysis performance.

III. WELL LOG FUZZY PREPROCESSING

In well log data analysis, only data preprocessing in the form of verifying the available core data will be discussed in this paper. The log analyst will base on some heuristic rules to prepare the training core data. Such rules are normally derived from past experience, established theories on the derivation of the petrophysics characteristics and the knowledge of the wells. This data preprocessing process, while in many cases is semi-automatic, is still a very time consuming task. The time required also depends on the amount of available core data. Obviously, more core data will require more time for preprocessing.

Before formulating the fuzzy preprocessing approach, an investigation on how the log analysts make use of their knowledge needs to be carried out. The steps are outlined as follows:

1. Taking the typical range for all input logs and output core data, adjust the plot for visualisation purpose.
2. Based on the known heuristic rules, carry out visual inspection on the plot. For example if the gamma ray is shown to be less than 20% of the range, the porosity can be expected to possess a high value.
3. Repeat Step 2 until most input logs have been verified with the output core. Note that the input logs are treated as independent of the other logs in the above steps.
4. If any input logs and output core data do not conform to the rules, they are either removed or adjusted manually. This depends entirely on the log analyst's experience.

If the amount of available input logs and training core data is huge, the first three steps mentioned above will be very time consuming. Beside, this method of data preprocessing is also subjected to human errors. If an intelligent and automatic technique can be used to replace the first three steps mentioned above, the log analyst can be free from these tedious tasks. At the same time, the possible occurrence of human errors could also be reduced.

Based on the steps described before, the fuzzy preprocessing technique is formulated as follow:

1. The log analyst will code some initial knowledge of the well log responses in the form of fuzzy rules. For example:

   a) If GR is LOW then KH is HIGH
   b) If RHOB is LOW then KH is HIGH

2. After all the initial rules have been set up, the range of the fuzzy memberships is determined. For the ease of coding the fuzzy rules, triangular membership function is recommended.
3. Based on the initial fuzzy rules, a more complex rules that link all the initial knowledge will be formed.

   For example:

   If GR is LOW and RHOB is LOW and KH is LOW then FALSE

4. The linked rules that are created will then be used to verify all the core data.
5. Any core data that are found to violate the heuristics rules will be discarded and are reported to the user.
6. After the fuzzy pre-processing process, only core data that conforms to the human heuristic rules will be fed into the BPNN model for training purpose.

IV. CASE STUDY AND DISCUSSIONS

A typical case study has been used to test the applicability of the proposed fuzzy preprocessing technique. In this case study, two wells that obtained from the same region are used. There are a total of 54 core data in Well A, and a total of 117 core data in Well B. The input logs that are available are: neutron (NPHI), sonic travel time (DT), bulk density (RHOB), and gamma ray (GR). The interested output petrophysical property is permeability (KH). In this case study, only Well A is used in the training process. Well B that is not used in the training process will serve as a benchmark to determine the accuracy of the prediction model.

An example of the initial knowledge that was acquired from the log analyst is shown in Fig. 1. Having set up all the individual heuristic rules, a set of complex fuzzy pre-processing rules that link all the initial rules are then set up. An example of these rules is shown in Fig. 2. Their corresponding membership functions are also shown in Fig. 3. It should be noted that the range used by KH is in logarithmic scale. For the ease of display, the x-axis for RHOB and KH has been scaled up by a factor of 10. The fuzzy memberships are labelled as L, M and H from left to right respectively in Fig. 3.

If GR is LOW then KH is HIGH
If RHOB is LOW then KH is HIGH
If NPHI is HIGH then KH is HIGH
If DT is HIGH then KH is HIGH

Fig. 1: The initial heuristic rules

If NPHI is H and DT is H and RHOB is L and GR is L and KH is L then False
If NPHI is L and DT is L and RHOB is H and GR is H and KH is H then False

Fig. 2: Fuzzy preprocessing rules

After the fuzzy preprocessing rules have been set up, they are used to verify the core data in Well A. It was found that there are a total of 11 core data that do not agree with the fuzzy preprocessing rules. In order to test the effect of the undesired core data has on the final BPNN model, two networks (BPNN A and B) are trained as shown in Table 1.

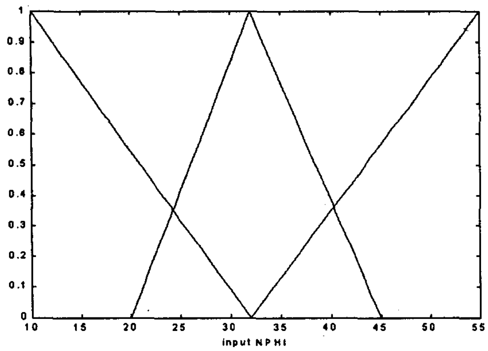

Fig. 3 (a): Fuzzy membership for input NPHI

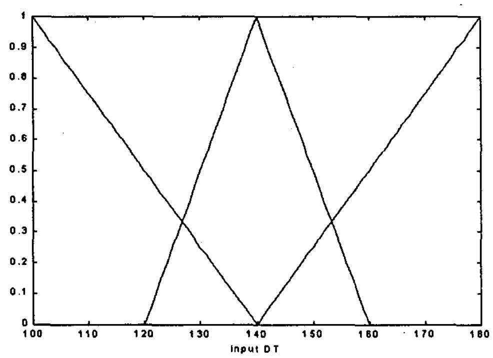

Fig. 3 (b): Fuzzy membership for input DT

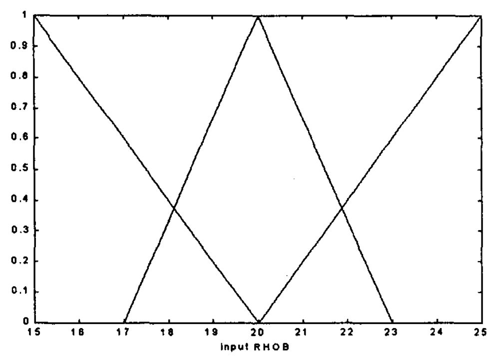

Figure 3 (c): Fuzzy membership for input RHOB

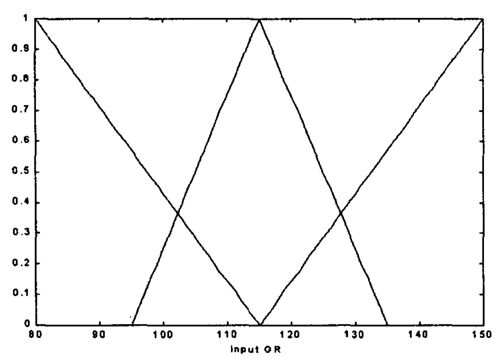

Fig. 3 (d): Fuzzy membership for input GR

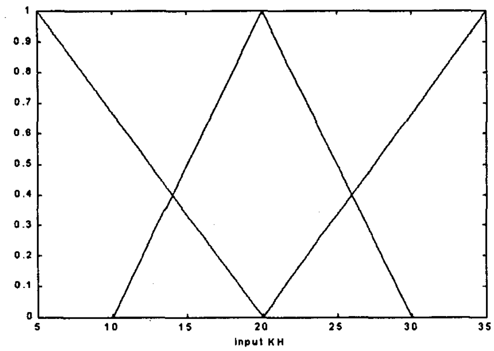

Fig. 3 (e): Fuzzy membership for output KH

Table 1: Comparison of two networks used in the case study

	BPNN A	BPNN B
With Fuzzy Preprocessing	No	Yes
No. of training data	54	43
No. of hidden units used	8	8
Training time	232 sec	206 sec.
Percent Similarity Coefficient of predicted KH in Well B as compared to its core data	52.641	82.671

From the user's viewpoint, it is desirable that the prediction model could generate accurate prediction on data that have not been used in the training process. The accuracy of the prediction of the properties of Well B is therefore of great interest. Two sets of predicted outputs have been generated in this case study: BPNN A that uses all training data, and BPNN B that only uses core data that are conformed to the heuristic rules derived from the log analyst's experience. The difference between the predicted outputs as compared to the actual core data is calculated from the *Percent Similarity Coefficient* expressed in (8). The results are shown in Table 1.

Percent Similarity Coefficient:

$$PSC = 200\frac{\sum_{i=1}^{P}\min(T_i, O_i)}{\sum_{i=1}^{P}(T_i + O_i)} \qquad (8)$$

The output plots of both the training well and testing well are shown in Fig. 4 and Fig. 5 respectively. The solid line (NOCHK) presents the output generated from the BPNN A, and the dotted line (CHK) presents the output generated from the BPNN B. The dots (KH) show the core data of the well. The last column (FUZZY) in the plot of Figure 4 indicates the result of the fuzzy preprocessing. Any value that falls in 0 indicates that there is a volitation of the fuzzy preprocessing rules. In BPNN B, these data have been discarded during training. From observing the plots, BPNN A performs well in the training well, but predict badly in the testing well between 2360 to 2380 metres. This indicates that the 11 noisy core data appeared in the training well have affected the final prediction model.

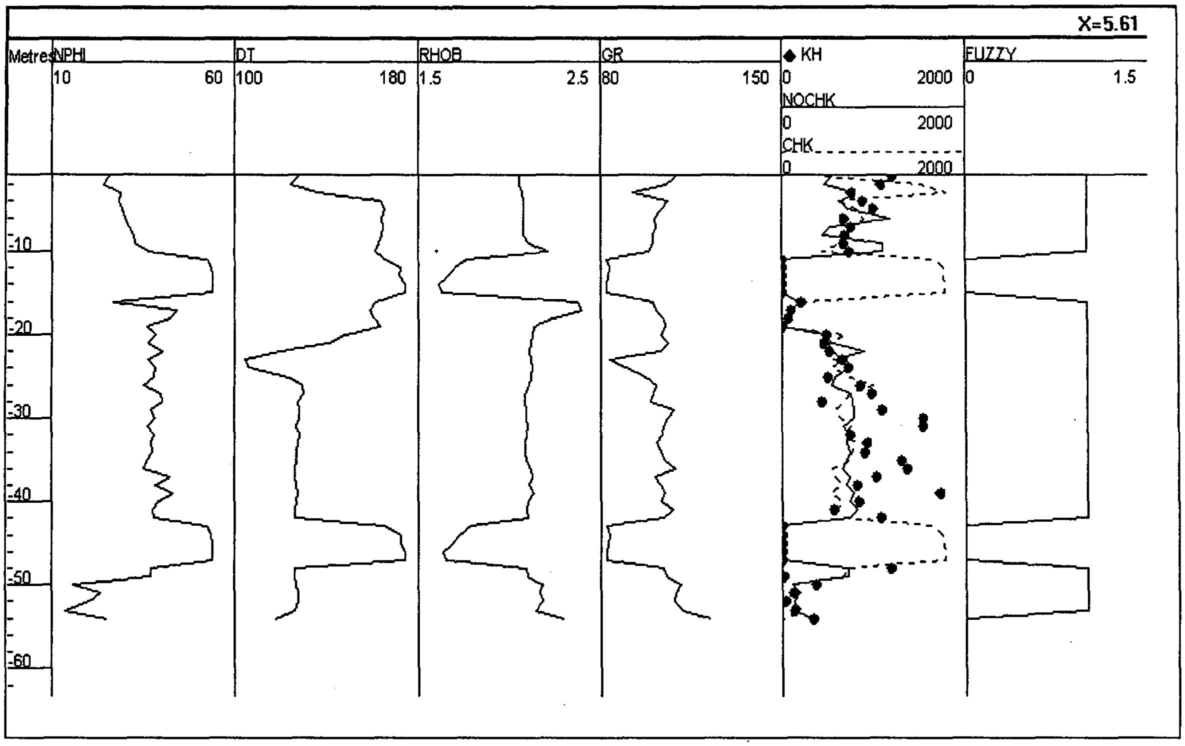

Fig. 4: **Graphical plot of the training Well A**

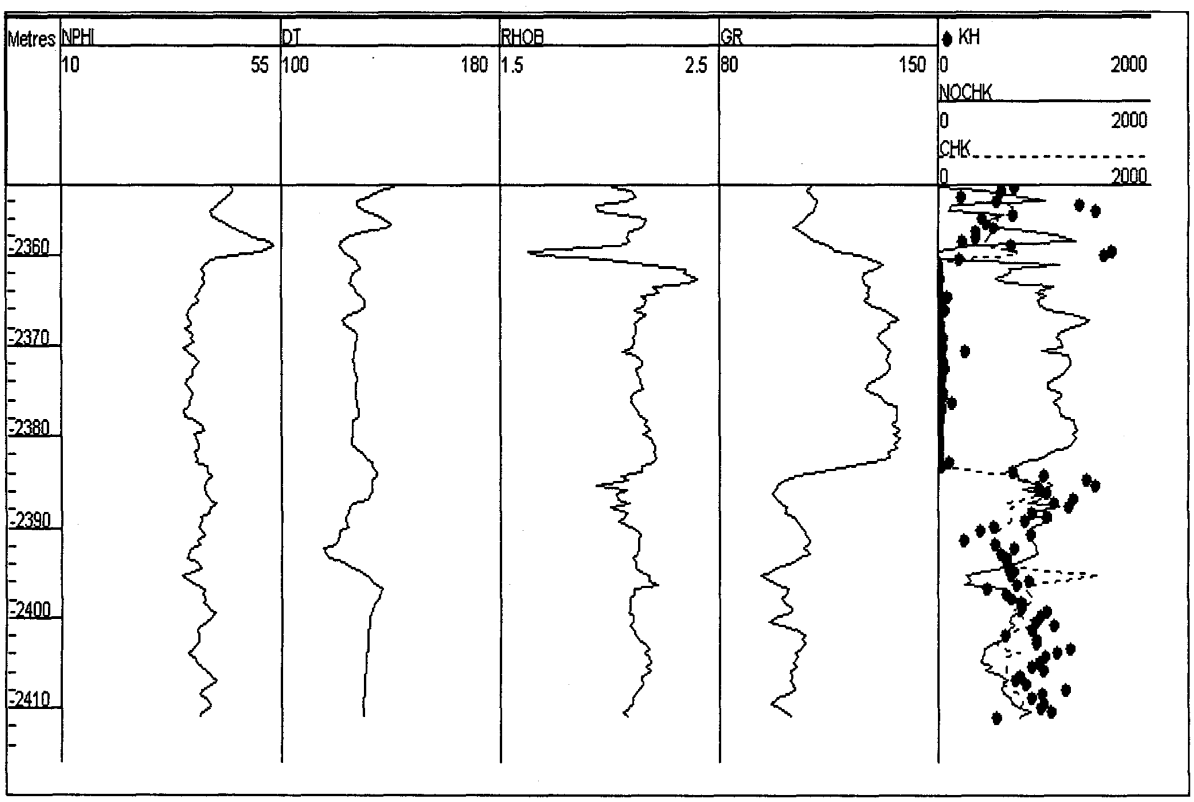

Fig. 5: **Graphical plot of the testing Well B**

V. CONCLUSION

This paper has taken a look at the effect of noise or unwanted training data has on a BPNN based well log interpretation model. This paper also highlighted the importance of preprocessing in determining the accuracy of the prediction model. Finally, the paper proposed a fuzzy preprocessing technique to improve the accuracy of the ANN well log interpretation model.

As the computational time for verifying the core data is small, the log analyst could be free from those long man-hours to check the log and core data manually. As fuzzy rules are expressed in linguistics terms, log analyst would not need much time in coding the rules. However, after the rules have been coded, they can be re-used in the analysis of other similar cases.

This paper also presented results from an experimental study showing that with the implementation of the fuzzy preprocessing technique, the prediction accuracy of the BPNN well log interpretation model could be increased. This new method has the potential to be a useful and important tool for professional well log analysts.

ACKNOWLEDGEMENT

This research was supported by a Small Grant from the Australian Research Council.

VI. REFERENCES

[1] D.V. Ellis, Well Logging for Earth Scientists, Elsevier Science Publishing Co. 1987.

[2] E.R. Crain, The Log Analysis Handbook Volume 1: Quantitative Log Analysis Methods, Penn Well Publishing Company. 1986.

[3] C.C. Fung, K.W. Wong, H. Eren, R. Charlebois, and H. Crocker, "Modular Artificial Neural Network for Prediction of Petrophysical Properties from Well Log Data," IEEE Trans. on Instrumentation and Measurement, vol. 46, no. 6, pp. 1295-1300, 1997.

[4] P.M. Wong, F.X. Jian, and I.J. Taggart, "A Critical Comparison of Neural Networks and Discriminant Analysis in Lithofacies, Porosity and Permeability Predictions", in Journal of Petroleum Geology, vol. 18(2), pp. 191 – 206, 1995.

[5] C.A., Goncalves, P.K., Harvey, and M.A., Lovell, "Application of a Multilayer Neural Network and Statistical Techniques in Formation Characterisation," SPWLA 36th Annual Logging Symposium, 1995.

[6] C.C. Fung, and K.W. Wong, "Petrophysical Properties Interpretation Modelling: An Integrated Artificial Neural Network Approach," International Journal of Systems Research and Information Science, vol. 8, pp. 203-220, 1999.

[7] H. Crocker, C.C. Fung, and K.W. Wong, "The STAG Oilfield Formation Evaluation: A Neural Network Approach", APPEA'99 Journal, Vol 39, Part 1, pp. 451-460, 1999.

[8] S. German, E. Beinenstock, and R. Doursat, "Neural Networks and the Bias/Variance dilemma", Neural Computation, 4, pp. 1-58, 1992.

[9] H. White, "Learning in Artificial Neural Networks: A Statistical Perspective", Neural Computation, 1, pp.425-464, 1989.

[10] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning Internal Representation by Error Propagation", Parallel Distributed Processing, vol. 1, Cambridge MA: MIT Press, pp. 318-362, 1986.