Heart Disease Detection

A Performance comparison using different Neural Network Algorithms.

Rajesh Thomas,

Dept. of Electrical Engineering,

FAMU-FSU College of Engineering,

2525 Pottsdamer Street,

Tallahassee, FL, USA.

rr15f@my.fsu.edu

Sourindu Chatterjee,

Dept. of Electrical Engineering,

FAMU-FSU College of Engineering,

2525 Pottsdamer Street,

Tallahassee, FL, USA.

sc15@my.fsu.edu

Dr. Simon Foo,

Dept. of Electrical Engineering,

FAMU-FSU College of Engineering,

2525 Pottsdamer Street,

Tallahassee, FL, USA.

sfoo@fsu.edu

*Abstract* — This project utilizes various Neural Network algorithms to predict the occurrence of Coronary Heart Disease (CHD) in a patient with relatively high accuracy. The data for training the networks is taken from the UCI database. This project also aims at comparing the performance of the various algorithms used, namely, K-Nearest Neighbor (KNN), Backpropagation, Radial Basis Network (RBN) and Decision Trees. Pre-processing is done using Principal Component Analysis (PCA) which enables feature reduction. The work done has led to conclusion that RBN provides the maximum accuracy for the UCI database.

*Index Terms* — Neural Network, K-Nearest Neighbor, Backpropagation, Radial Basis Network, Decision Tree, Coronary Heart Disease.

# Introduction

All manuscripts must be in English. These guidelines include complete descriptions of the fonts, spacing, and related information for producing your proceedings manuscripts. Please follow them and if you have any questions, direct them to the production editor in charge of your proceedings (see author-kit message for contact info).

This template provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. PLEASE DO NOT RE-ADJUST THESE MARGINS. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

# Pre-Processing

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts. True Type 1 or Open Type fonts are required. Please embed all fonts, in particular symbol fonts, as well, for math, etc.

# Algorithms

This project compares the performance of four different algorithms for detecting coronary heart disease. The pre-processed data is used as input data. The four algorithms used in this work are K-Nearest Algorithm, Backpropagation, Radial Basis Network and Decision Trees. They were chosen because these algorithms are quite prominent in machine leaning and are relatively easy to implement using powerful tools like MatLab.

## K-Nearest Neighbour

K-Nearest Neighbor (KNN) is quite an easy to implement algorithm which is also conceptually easy to understand. KNN works by checking the ‘K’ closest features to the new input. The feature with the maximum frequency is selected. Thus KNN can account for local variations in features. There are different methods to find the closest features. The more common ones are using Euclidean distance, Hamming distance, etc. The accuracy of the output also depends on the value of K. K is chosen so that it should neither be too big nor too small.

In two dimensions this can be easily imagined by using a dataset with two input features. One of the features is taken along the X-axis and the other along the Y-axis. A circle is drawn around a new data point so that K values are inside the circle. The class for the new point is chosen as the class that has the greatest frequency among the K values.

Using MatLab, optimization can be done so that the best value of K and the best distance metric can be found for the given training data. The graphs given in [Fig. 1.](#Figure1) show this.



Fig. 1. KNN Optimization Graphs

## Backpropagation

Backpropagation (also known as Multi-Layer Feed-forward Network) is a very versatile algorithm which can adapt to different scenarios, aptly earning the name ‘the workhorse of learning in neural networks’. Backpropagation works by assigning weights to the input vector and the hidden neurons. The weights are changed by calculating the gradient of a loss function with respect to all the weights in the network. The gradient is fed to an optimization function whose output is used to update the weights. Since backpropagation initially needs some outputs for training (to calculate the weights), it is considered as a supervised training method. The figures obtained for training from MatLab is given in [Fig. 2](#Figure2).

For this project there are a total of ten input features which are reduced to three using Principal Component Analysis (PCA). Training is done on the inputs and the accuracy is tested using new input data. Optimization is done to find the minimum number of hidden neurons required.

The output is calculated using the equation

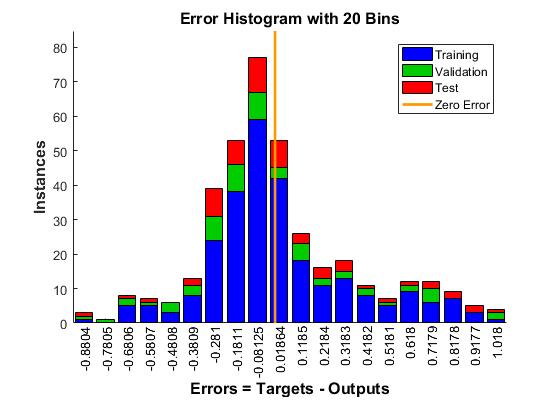
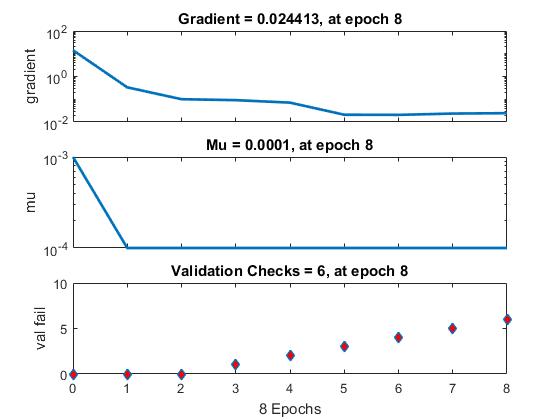
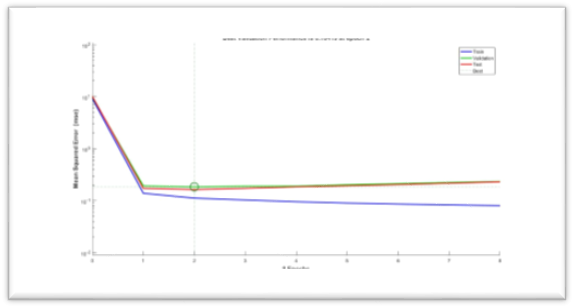


Fig. 2. Backpropagation Optimization Graphs

## Radial Basis Network

Radial basis network (RBN) works using the concept of Radial Basis Functions (RBF). RBN works by approximating functions using the given data by iterative training. The error on the training data can be reduced to values as small as desired. New neurons can be added until the means square error (MSE) goal is achieved. Thus for a given number of discrete data inputs, a function is made in such a way that it passes through the given inputs as closely as possible.

The function approximation is usually made using the form given in the following equation:



Where φ is a radial basis function which include Gaussian, Multiquadric, Polyharmonic spline, etc., and wi represents the weights. While implementing as an RBN, φ takes the role of activation function. The weights are updated by finding the error of y(x). This is obtained by differentiating y(x) with respect to the weights.

The features extracted from PCA is used for training RBN. The training graph obtained is given below in Fig. 3.

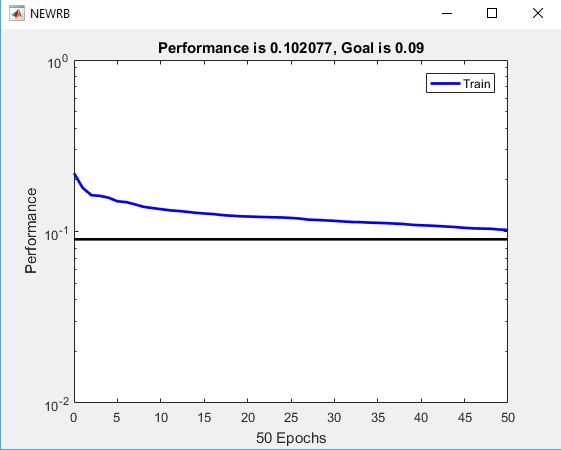


Fig. 3. Error during RBN training.

## Decision Tree

Decision trees are decision support tools which has a flowchart like structure. The decisions are taken using the relevance of the input features. The relevance is found by calculating the information entropy contained in the features which is often done using the Gini index. The feature with the maximum gain in entropy is considered for splitting the tree.

Training is done by growing the tree based on the entropy of the input data. After the tree has been made, testing is done on new inputs. The optimization graphs are shown below in Fig. 4 and the decision tree in Fig. 5.



Fig. 4. Optimization graphs for decision tree.

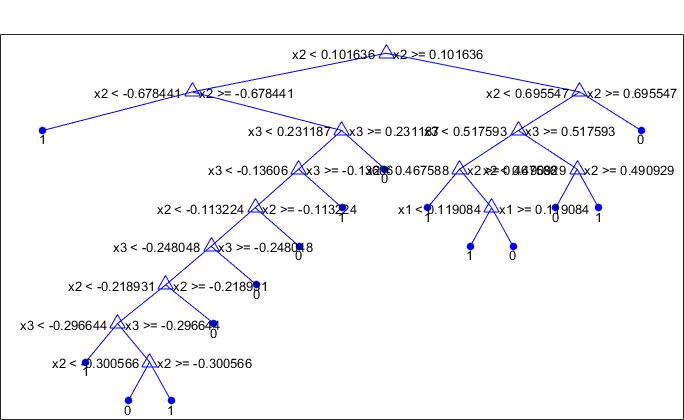


Fig. 5. A decision tree grown from the PCA inputs.

# Results

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left ofq the MS Word Formatting toolbar.

## Authors and Affiliations

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

### For Author/s of Only One Affiliation (Heading 3): To change the default, adjust the template as follows.

#### Selection (Heading 4): Highlight all author and affiliation lines.

#### Change Number of Columns: Select Format > Columns >Presets > One Column.

#### Deletion: Delete the author and affiliation lines for the second affiliation.

### For Authors of More than Two Affiliations: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change Number of Columns: Select Format > Columns > Presets > One Column.

#### Highlight Author and Affiliation Lines of Affiliation 1 and Copy this Selection.

#### Formatting: Insert one hard return immediately after the last character of the last affiliation line. Then paste down the copy of affiliation 1. Repeat as necessary for each additional affiliation.

#### Reassign Number of Columns: Place your cursor to the right of the last character of the last affiliation line of an even numbered affiliation (e.g., if there are five affiliations, place your cursor at end of fourth affiliation). Drag the cursor up to highlight all of the above author and affiliation lines. Go to Format > Columns and select “2 Columns”. If you have an odd number of affiliations, the final affiliation will be centered on the page; all previous will be in two columns.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is *“Heading 5”*. Use *“figure caption”* for your Figure captions, and *“table head”* for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named *“Heading 1”*, “*Heading 2”*, *“Heading 3”*, and *“Heading 4”* are prescribed.

## Figures and Tables

Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table captions should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1” in the text, and “Figure 1” at the beginning of a sentence.

Use 8 point Times New Roman for figure labels. Use words rather than symbols or abbreviations when writing figure-axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”.

If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

## Footnotes

Use footnotes sparingly (or not at all) and place them at the bottom of the column on the page on which they are referenced. Use Times 8-point type, single-spaced.

To help your readers, avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).

Number footnotes separately from reference numbers, and in superscripts. Do not put footnotes in the reference list. Use letters for table footnotes.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
|  | Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

a. Sample of a table footnote. *(table footnote)*

1. Example of a figure caption. *(figure caption)*

We suggest that you use a text box to insert a graphic (ideally 300 dpi, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on Example of a figure caption. *(figure caption)* your frame, use the MSWord pull-down menu, select Format > Borders and Shading > Select ”None”.

# Conclusion

You must submit the IEEE Electronic Copyright Form (ECF) as described in your author-kit message. THIS FORM MUST BE SUBMITTED IN ORDER TO PUBLISH YOUR PAPER.

# Acknowledgment

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression, “One of us (R. B. G.) thanks . . .” Instead, try   
“R. B. G. thanks”. Put applicable sponsor acknowledgments here; DO NOT place them on the first page of your paper or as a footnote.

# References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example: [1]. Where appropriate, include the name(s) of editors of referenced books. The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in “[3]”—do not use “Ref. [3]” or “reference [3]”. Do not use reference citations as nouns of a sentence (e.g., not: “as the writer explains in [1]”).

Unless there are six authors or more give all authors’ names and do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.