ECE 511: Neural Network Based Pattern Recognition and Classification

Will Chu
Roger Lai
Sherry Lai
Jene Park
Bill Pottle
December 19, 2001

EXECUTIVE SUMMARY

Neural Networks can provide a powerful method for classifying data. There is a strong need for automated data classification systems in the medical field. The ECG reading can record a number of distinct abnormal heartbeats after which a trained doctor can pick them out. However, if the ECG systems can monitor a patient for 24 or more hours, then they can provide much more useful information. Doctors and clinicians simply do not have the time to go through tremendous amounts of patient data. Thus, an automated system must be used.

Our company developed a neural network system capable of correctly classifying over 86% of the heartbeats that it was presented with. The system consists of a 3-layer network with one input layer, one hidden layer, and one output layer. The input layer is 300 nodes big and consists of values for the windowed data points. The window is approximately one second long, which allows for one heartbeat. The peak of the QRS wave is aligned to the 100th node. In our network, the number of nodes in the middle layer varied between 10 and 15 nodes, and the output layer consisted of five nodes, labeled L, A, V, R and N. "L" represents a *Left Bundle Branch Block*, "R" represents a *Right Bundle Branch Block*, "A" represents a *Premature Atrial Contraction (PAC)*, "V" represents a *Premature Ventricular Contraction (PVC)*, and "N" represents a normal beat. The highest output node determined what classification the network was assigned to for a particular beat.

We developed our system by training networks under different training conditions and by batch testing the resultant networks against a set of windowed patient records annotated by Massachusetts Institute of Technology/Beth-Israel Hospital (MIT/BIH) cardiologists. We developed an F value, f(1.00*Correct-0.60*FNeg-0.20*Unclass-0.30*FPos), that weighted errors differently. This was our main criteria in ranking the networks. Indicating to the patient that he is fine when he is actually sick (False Negative) is much worse than telling a healthy patient that he may have a problem (False Positive). The False Positive case will at least be followed up and later determined that the condition is not serious. We also tested networks for their ability to pick out *Premature Atrial Contractions (PAC)* and *Premature Ventricular Contractions (PVC)* from the set. Thus, another key parameter was the percentage of A or V beats that the network correctly identified.

As such, our company has two networks. Our main network works well on the average, but has two classification errors. It sometimes classifies N beats as A, and L beats as V. These waveforms are very similar visually (cf. 10 and 11). Although the main network performed better on average and had an F value of 75.05, it caught only 70.7% of all the abnormal A and V beats. Thus, we developed another network specifically for A and V detection, and this network caught 87.03% of the A and V beats, outperforming the competitor's network. This network was called the A/V network and had an F value of 63.85. The main network was trained on the "TrainSB.bin" dataset, while the A/V network was trained on the "TrainD.bin" dataset.

NEURAL NETWORKS

The Biological Foundation of Neural Networks

The human nervous system consists of an extremely large number of nerve cells, or neurons, which operate in parallel to process various types of information.

As the central station of the nervous system, the brain is made of cells called neurons which communicate with each other at places called synapses. A synapse is a functional contact between two neurons. There are about 100 billion neurons in the human brain and each has about 10,000 contacts with other neurons. The number of synapses in the human brain is about 10^{15}. An interesting fact is that the neurons of one human cerebral cortex would reach over 250,000 miles if placed end to end.

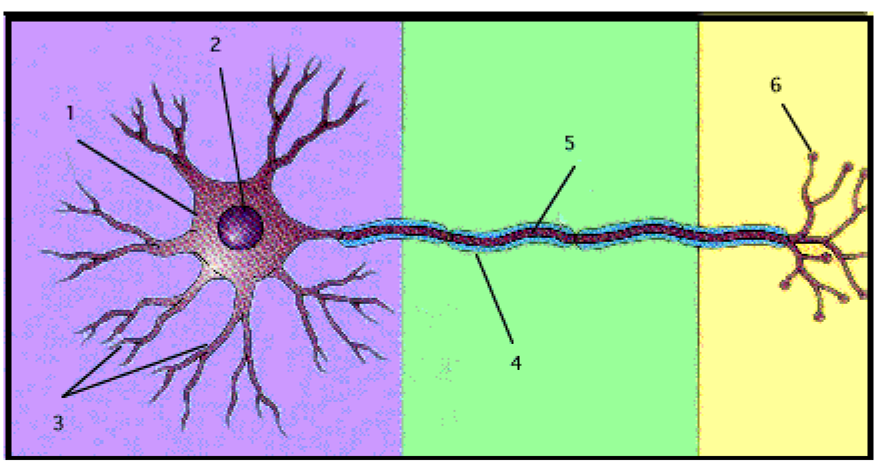


Fig. 1. A Neuron: 1. Cell Body (Soma), 2. Nucleus, 3. Dendrites, 4. Axon,
5. Myelin Sheath, 6. Axon Terminals

The transmission of signals from one neuron to another at synapses is a complex chemical process. Specific transmitter substances are released from the sending end of the junction. The effect is to raise and lower the electrical potential inside the body of the receiving cell. If the potential reaches a threshold, a pulse is sent down the axon.

Artificial Neural Networks

Artificial neural networks (ANN's) are collections of mathematical models that emulate some of the observed properties of biological nervous systems and draw on the analogies of adaptive biological learning.

They are composed of a large number of highly interconnected processing elements (nodes) that are analogous to neurons and are tied together with weighted connections that are analogous to synapses.

Learning in NNs is based on that of biological systems, which involves adjustments to the synaptic connections that exist between the neurons. Learning typically occurs by example through training, or exposure to a set of input/output data, where the training algorithm iteratively adjusts the connection weights. These connection weights store the knowledge necessary to solve specific problems.

A powerful property of neural networks is that they possess the ability to generalize in making decisions about imprecise input data.

A Simplified Mathematical Model of the Neuron

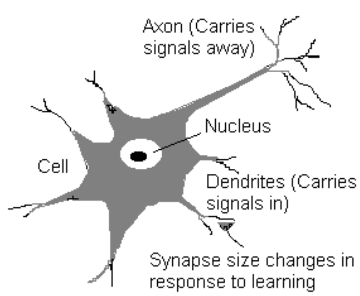


Fig. 2. A simple neuron cell

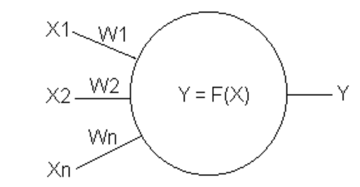


Fig. 3 Schematic of a neuron

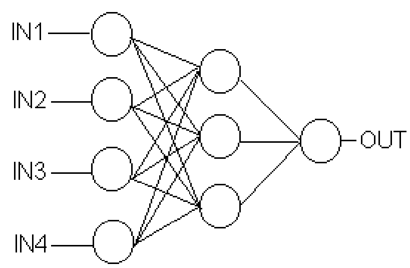


Fig.4 Feed Forward Neural network

In a neural network, the effects of the synapses are represented by "weights" which effectively models the nonlinear characteristics exhibited by neurons represented by a transfer function. The neuron impulse is then computed as the weighted sum of the input signals, transformed by the transfer function. The learning capability of an artificial neuron is achieved by adjusting the weights in accordance to the chosen learning algorithm. Examples of learning algorithms include the Delta Rule, Hebb's Rule, Hopfield Law, and Kohonen's Learning Law. Neural networks offer ideal solutions to a variety of classification problems such as speech, character and signal recognition, functional prediction and system modeling, and control problems (where the network learns the control function). The advantages of NNs are (1) resilience against distortions in the input data and its capability of learning in areas of application such as prediction - using input values to predict some output, (2) classification - using input values to determine the classification, (3) data conceptualization - analyzing the inputs so that grouping relationships can be inferred, (4) data association – classifying but also recognizing data that contains errors, and (5) data filtering.

HEART ABNORMALITIES

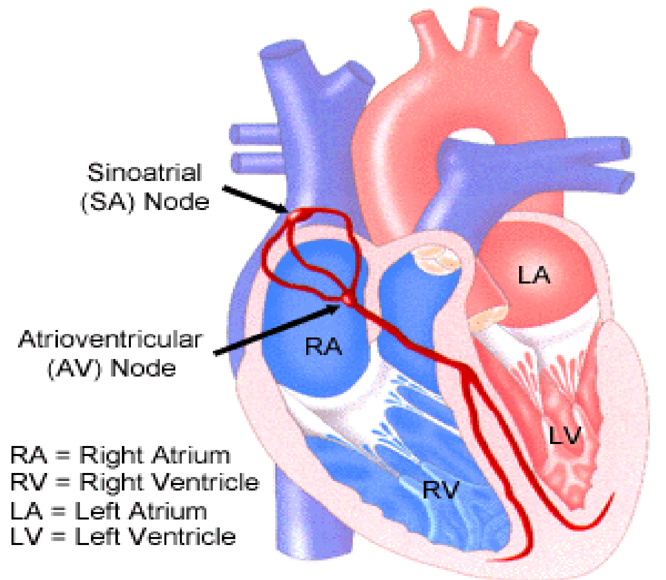


Fig 5. Electrical System of the Heart

N – *Normal Sinus Rhythm*

Normally, the heart rhythm is controlled by electrical impulses arising in the sinus node portion of the right atrium. The impulses travel through both atria, and then enter the left and right ventricles through the AV node and the His bundle. With a normal pattern, the atria contract first, emptying their blood into the ventricles, and then the ventricles to contract.

Premature Beats are not a main concern, as they are common in normal individuals. Their presence alone does not require treatment unless they cause intolerable symptoms. Very often, a few changes in lifestyle, such as reducing stress or avoiding caffeine, will be sufficient to regulate the heartbeat.

V – *Premature Ventricular Contraction*

PVCs can simply be a normal variation in the heart's healthy operation or can mean that there is a serious underlying heart problem. Thus it is important for a network to distinguish PVCs from other types of beats.

A – *Premature Atrial Contraction*

PACs originate within the atrial myocardium but outside the sinoatrial node. The PACs occur before the next expected sinus discharge. Premature beats are very common in normal children and teenagers – most have them at some time. Usually no cause can be found and no special treatment is needed. The premature beats may disappear later.

Electrical impulses normally travel down both the right and left bundles at the same time, after which the left and right ventricles contract simultaneously. Bundle bunch blocking (BBB) is caused by a disruption in the conduction of the electrical impulses. BBB is quite common and occurs in a variety of medical conditions. BBB can be diagnosed from the ECG, where the QRS complex is significantly wider than normal.

R – *Right Bundle Branch Block*

For this condition, the right bundle no longer conducts electricity. An electrical impulse leaves the bundle of His and enters only the left bundle branch and is carried only to the left ventricle. Only then does the electrical impulse makes its way from the left ventricle to the right ventricle. As a result, the two ventricles no longer contract at the same time. However, RBBB commonly occurs in normal, healthy individuals, and screening exams therefore often turn up with no medical problems.

L – *Left Bundle Branch Block*

The left bundle branch block is similar to an RBBB, except now the left bundle no longer conducts electricity. However, LBBB usually indicates underlying cardiac pathology. It is seen in dilated cardiomyopathy, hypertrophic cardiomyopathy, hypertension, aortic valve disease, coronary artery disease, and various other cardiac conditions. LBBB may also occur in healthy people, but a thorough search, as opposed to simple screening, should be conducted for underlying cardiac problems.

TRAINING METHODS

To carry out the Neural Network Training, we used NNCAD for Windows 3.3. For this lab, the feed-forward neural networks (FFNN) were employed.

Default training parameters are as follows:

1) Number of layers: 1
2) **Number of nodes in 1st layer: 15**
3) Number of nodes in 2nd layer: 0
4) Dimension of input: 300
5) **Maximum number of iterations: 100**
6) **Initial learning rate: gamma = 0.000100**
7) **Error goal: 0.010000**
8) Random seed for initial weights: 16
9) Ad hoc error ratio: er = 1.0005000
10) **Ad hoc learning rate decrease: dm = 0.700000**
11) **Ad hoc learning rate increase: im = 1.050000**
12) F_{min} gamma termination tolerance: xtermin = 0.000100
13) F_{min} MSE termination tolerance: ftermin = 0.000100
14) Folder\filename of training set: defaultf\TrainSB.bin
15) Folder\filename of generalization set: defaultf\Gen1.bin
16) Scale training set vectors by 10.000000.
17) Evaluate the generalization set every 50 iterations.

18) **Name of output folder\filename: testf\test**
19) Autosave every 25 iterations.
20) Do not append iteration number to autosave filenames.

For the purposes and simplifications of the lab, our group only dealt with specific variances of parameters to obtain the highest possible f-value – Number of Nodes, Maximum Number of Iterations, and Gamma.

F-value is a direct measurement of network performance. The number takes into account four different types of classifications – (1) *Correct Classification* (2) *False Positive* (3) *Unclassified* (4) *False Negative*, in the order of best to worst cases. *Correct Classification* was denoted with a value of 1.0. We considered the weighting factors for *False Positive* and *Unclassified* as relatively close and therefore were given the values of 0.2 and 0.3 respectively. Finally, the worst case scenario of *False Negative* was denoted by a weighting factor of 0.6. *Correct classification* of a heart condition is the case in which the correct diagnosis is made. A *false positive* is a diagnosis made for a condition that is actually not there. A *false negative* is when a wrong or no diagnosis is made for a condition that is actually present. Finally, *unclassified* cases are when a firm diagnosis cannot be made and further analysis is needed. *Correct classification* was the best case scenario, *false positive* second best, then *unclassified* and finally *false negative*. Rating between *false positive* and *unclassified* was more difficult. In the event of a *false positive* or an *unclassified*, the doctor will have to personally examine the waveforms, so it is better to have the network classify a waveform. Therefore, the f-value can be calculated using the weighing factors for all four types of classifications.

$$\mathbf{F}\boldsymbol{val} = \boldsymbol{f}\,\boldsymbol{(}1.00\times\boldsymbol{Correct} - 0.60\times\boldsymbol{FNeg} - 0.3\times\boldsymbol{Unclass} - 0.2\times\boldsymbol{FPos}\boldsymbol{)}$$

<u>Explanation of Parameters</u>

Number of nodes in the first layer of FFNN

Single layer FFNNs were investigated in this lab. The neural network is composed of many highly interconnected processing elements (nodes). A classic representation of the neural network includes - inputs (x_n), weights between input and nodes (W), α (between nodes and outputs), and outputs (y_n). Thus, the output is a direct function of input.

Maximum number of iterations

A weight is a specific neural pattern from the input (x) to its neuron (τ). Each input has a specific weight for its specified node (neuron). All weights are present between inputs and nodes to maximize interconnections between nodes. Adjustments to the synaptic connections that exist between the neurons will aid in learning about the biological system. Fortunately, the training sets can iteratively adjust the connection weights.

Gamma

This parameter determines how far the training set will adjust the weights and thresholds for each iteration.

<u>How values were chosen</u>

On the evening of December 8th, 2001, our team members ran five different training runs. To verify that data was properly being stored, we initially ran the first training run up to 10,000 iterations.

Number of Nodes

We chose from 10-13 nodes on each of our runs because increasing beyond a certain number of nodes for each training set would no longer enhance the performance of the system after a certain point. Performance may be improved if the number of iterations is increased significantly. However, as we note later, increasing the number of maximum iterations will affect the accuracy of interpreting patient data. If we were to plot the network performance as a function of a number of nodes, we would see that initially as the number of nodes is increased, performance highly increases. However, after the maximum threshold of nodes is reached, the performance will reach a plateau. In lecture, it was given that the number of nodes at which the performance begins to plateau is around 12. Therefore, we varied the number of nodes slightly below and above this threshold value.

Maximum number of iterations

As the number of iterations increased, f-value outputs also increased. Thus, the run with our highest number of iterations resulted in a high f-value. However, there is a threshold value for the number of iterations. With too many iterations, the set will be highly trained, and the f-values will be maximum. However, applying patient data to these networks will actually create many false classifications because the network is over-trained. Our group did not reach this limit.

Gamma

The default gamma is set to 0.0001. A gamma that is too small may lead to minimal training progress. Also, the system may get trapped in a local minimum error. However, a gamma that is set too large may result in unstable updates and less refinement of final weight values.

The following table summarizes the results and what parameters were varied.

Training Set	Number of nodes	Gamma	Maximum number of iterations
Default Training Set	15	0.000100	100
Training Set SB	12	0.000100	10,000
Training Set SB	12	0.0000500	75,000
Training Set SB	11	0.000200	50,000
Training Set SB	13	0.000100	100,000
Training Set SB*	**10**	**0.000100**	**150,000**

Table 1. Summary of parameters for each training run. Training Run 5 resulted in highest f-value

For the A/V classification analysis, we ran Training Sets A, B, C, and D with 12 nodes, gamma set a 0.0001, maximum number of iterations set at 10,000.

	Number of Nodes	Gamma	F-Value
Training Set A	12	0.000100	65.85
Training Set B	12	0.000100	67.36
Training Set C	12	0.000100	45.02
Training Set D	12	0.000100	63.85

Table 2. Summary of Training Networks on Different Training Sets. Training Set B had the highest f-value but poor classification of A/V vs. N/R/L. Training Set D had an overall high f-value and high percentage of correct classification between A's and V's.

TRAINING RESULTS

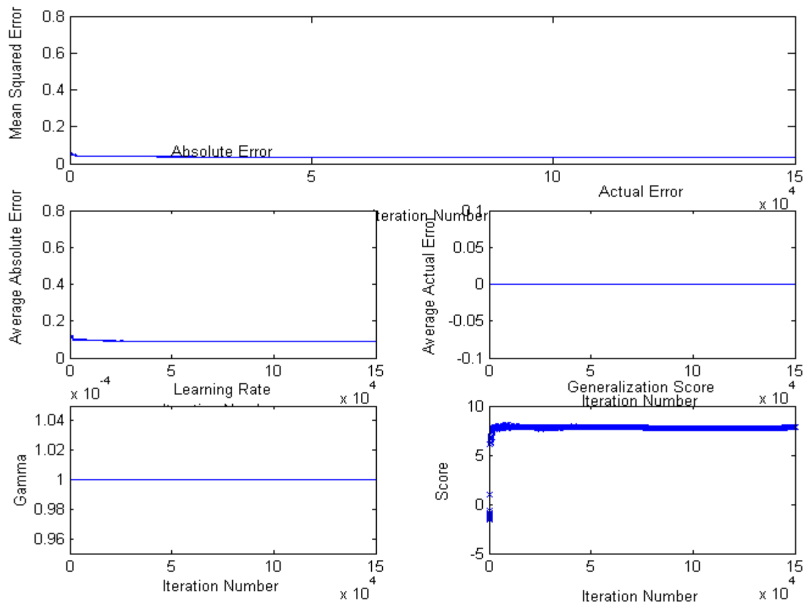


Fig. 6. Error Statistics in Training a Network

When training a neural network, the main point of interest is the average squared training error, χ^2.

$$\chi^2 = 1/(2 \bullet N_p) \bullet \sum_{p=1}^{Np} \sum_i \quad (t_p^i - o_p^i)^2$$

where N_p is the number of training patterns, t_p^i is the desired target value, for pattern *p* and output neuron *i*, and o_p^i the true output value from output neuron *i* for the same input pattern *p*.

1) The mean squared error with respect to iteration number decreases but eventually attains a steady state of slightly above 0. It is thus assumed that any errors that exist past the iteration number 3×10^4 are inherent in the system (between 0 and less than 0.1). The errors decreased with each iteration number and stabilized during the time of the number of iterations.
2) The actual error remains at 0 without averaging the entire set.
3) Gamma was kept as 0.0001, a small value throughout the set of iterations to allow for flexibility of the network to generalize new sets of inputs.
4) The generalization score appear to level out at 7 and increases at the very end. This is an indication that the network still has flexibility to generalize a new set of data even

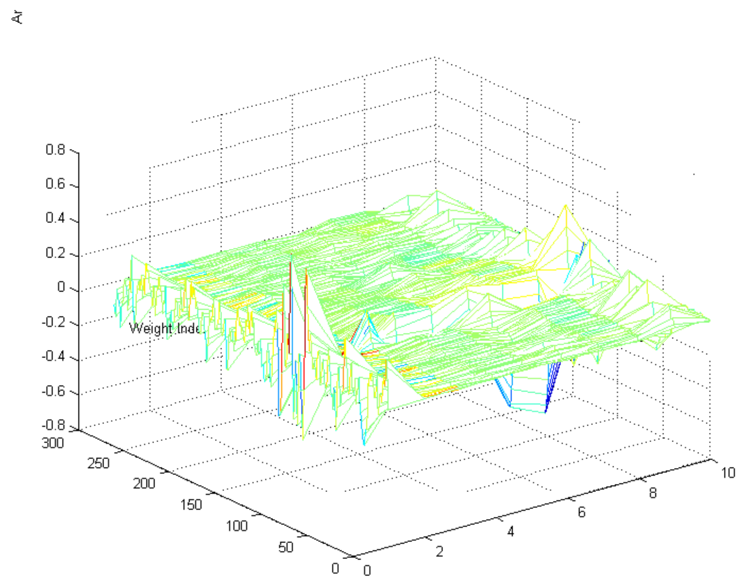


Fig. 7. Weight Index on a hidden layer consisting of 10 nodes after the 15,000 iterations

With a sample rate of 300 Hz, mapped over 10 nodes in the hidden layer, one notes the weight indices over each node. Apparently the weights do not change significantly about 0, with most activity occurring at the 100th sample. This is due to the fact that the sample is aligned to the QRS at the 100th sample.

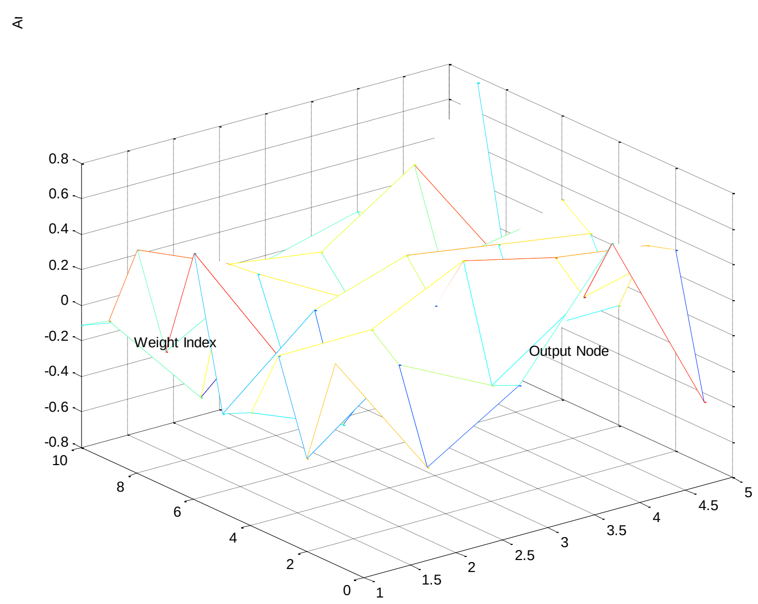


Fig 8. Weight Index at the Output Layer of 5 nodes

The above illustrates a similar surface plot of the amplitude weight indices from the hidden layer of 10 nodes to an output layer of 5 nodes.

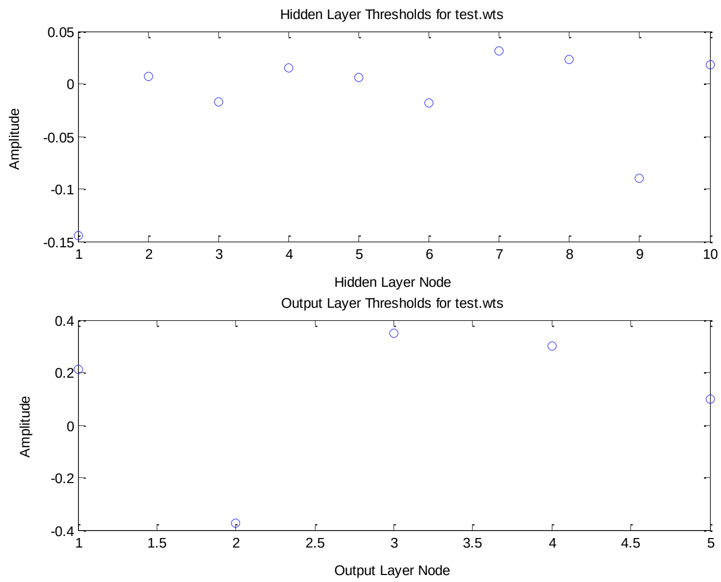


Fig. 9. Amplitude of Weights at the Hidden and Output Layers

NETWORK PERFORMANCE

Overall our network performed well. For all frames, it correctly classified 86.49% of the beats, gave false negatives 12.49% of the time, and gave false positives 3.21% of the time. The network failed to classify 1.02% of all beats. The formal statistics grid is given below:

Network Classification of Frames (StatsGrid)

	L	R	A	V	N	unclassified
L	6843	49	11	978	162	28
R	280	4496	39	12	170	37
A	72	73	104	44	130	32
V	217	29	35	1619	143	51
N	422	251	609	216	14236	173
O	325	90	74	344	2053	20

Table 3. Network Classification of Frames

The performance of the network also varied considerably depending on the type of beat it was trying to detect. It performed very well on L, R and N beats, marginally well on V beats, and poorly on A beats, classifying only 22.86 % of the A beats correctly. A beats were definitely a minority in the composite testing set. This caused the A beat misclassification to not matter as much in overall network performance. The results for each beat type are given below:

Beat TYPE	CORRECT	F/NEG.	F/POS	UNCLASS	TOTAL
L #	6843	1200	991	28	8071
L %	84.79%	14.87%	4.22%	0.35 %	
R #	4496	501	402	37	5032
R %	89.31%	9.95	1.52%	0.74%	
A #	104	319	455	32	694
A %	22.86	70.11%	2.23%	7.03%	
V #	1619	424	1250	51	2094
V %	77.32%	20.25%	4.24%	2.44%	
N #	14236	1498	173	605	15907
N %	89.5%	9.42%	3.86%	1.09%	

Table 4. Network Classification of Frames

Another important factor is how the network performed on normal vs. abnormal beats. Of the 15654 beats classified as L, A, V or R, our network classified the beat as normal 605 times (3.8%). This is critical because if a patient has an abnormal beat classified as normal by the network, the doctor may tell the patient they are fine and send them home. The results are broken down by normal and abnormal beats below:

Network Classification of Normal Frames (StatsGrid)

	Abnormal	Normal	Unclassified
Abnormal	14901	605	148
Normal	1498	14236	173
Other	833	2053	20

Table 5. Network Classification of Frames for Normal and Abnormal Beats

Since premature beats may signal some underlying heart condition, knowing when a beat is an A or a V is important as well. Our network performed fairly well on classifying A or V beats, catching 70.7% of them as such. Again, here the low amount of A values helped our classification because we performed well on V beats. While we beat most of the competitor's networks on overall classification (F value) we lost out on A or V classification. Another important number is the amount of LRNO beats that were classified as A or V. Here, our network classified 7.75% of the annotated LRNO beats as either A or V. However, our A/V network that was trained to maximize detection of A or V beats caught 87.03%, thereby outperforming the competitor's results.

Network	F value	Percentage of A or V beats identified as A or V
Competitor 8 node	60.67	86.65 %
Competitor 12 node	66.85	85.7 %
Competitor 20 node	77.53	84.11 %
Group Network	75.05	70.7 %
A/V Network	63.85	87.03 %

Table 6. Classification of A's or V's <u>Vs.</u> Other

Our network was hindered by two main problems: classifying N beats as A and classifying L beats as V. However, the beats have very similar morphologies. Below are images of all the N beats that were classified as A and all the true N beats. The beats visually look similar so misclassification does not represent a major flaw in the network.

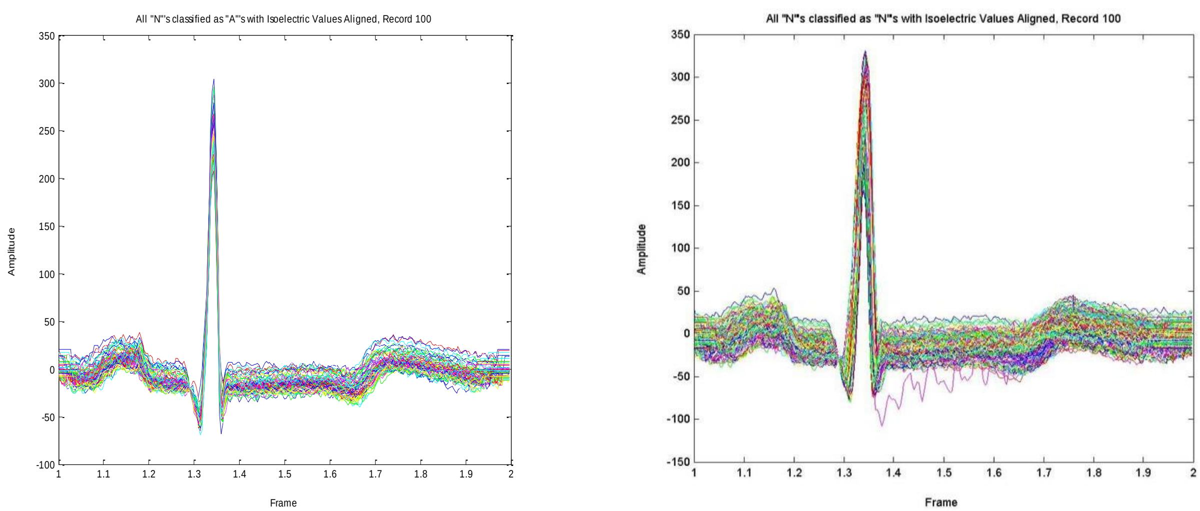
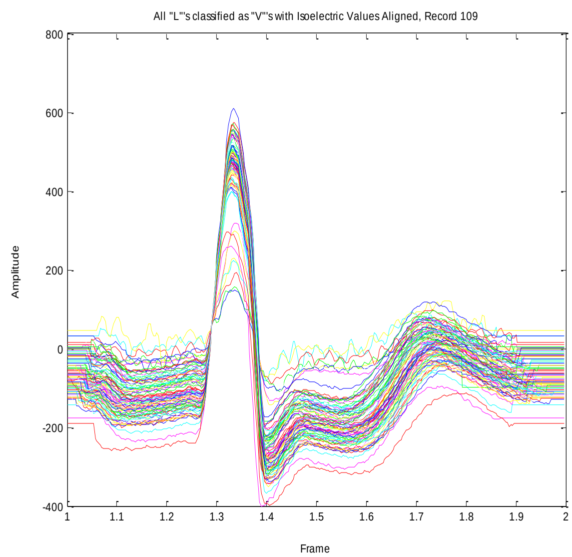


Fig. 10. Atrial Beat Misclassification – Record 100 had many A beats classified as N's. Comparison between A beats classified as N's, N's classified as A's, and N's classified as N's.



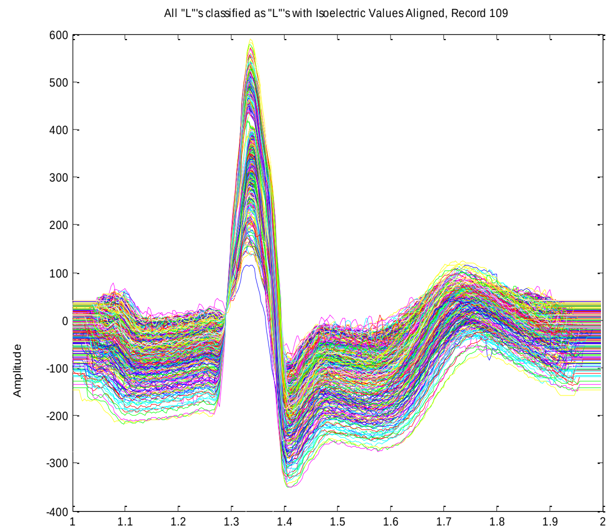


Fig.11. Ventricle Beat Misclassification. Record 109 had many L beats classified as V's. Comparison between V beats classified as L's, L's classified as V's, and L's classified as L's.

CONCLUSION

The analysis above illustrates the capabilities of our training network with respect to the those neural networks on the market. With an ability to detect A's or V's at 70.7 % for our main network, and an A/V network ability to determine 87.03% of A's or V's, our design is evidently comparable to the A/V detection of a market neural network for heartbeats (whose A/V detection statistics maintained a 85.48% percent on average for 3 different market neural networks). The importance of detecting premature beats, characteristic of such atrial or ventricular arrhythmias is thus fulfilled by this relatively high value. The ability of our network to detect premature heartbeats even more accurately would entail not simply training for the shape of the beat but rather run training sets with time variations in the waveforms. Furthermore, in addition to detecting premature beats, it remains equally as important to exhibit some capability of determining whether or not it originates from the atrial or ventricular portion of the heart. Judging from our classification results, a mere 4.01 % of V's were misclassified as A's, while 1.37 % of A's were misclassified as V's. This statistic bodes well for the application of our network for detailed diagnosis of premature heartbeats.

REFERENCES

http://sunflower.singnet.com.sg/~midaz/Intronn.htm
http://www.cs.stir.ac.uk/~lss/NNIntro/InvSlides.html#what
http://www.statsoftinc.com/textbook/stneunet.html#intro