**Abstract**

The challenge of diagnosing cancer is that no single test can accurately succeed. Diagnostic testing is essential to evaluate the health of an individual and determine whether the individual has cancer. Diagnostic imaging is a useful technique to produce an internal picture of the body for analyzing structure. However, medical professionals are required to successful analyze the images and determine whether the individual has cancer. Applying artificial neural networks to this problem makes the analysis more efficient while minimizing error in diagnosis.

The purpose of the project is to implement a successful neural network with backpropagation to analyze a breast cancer numerical and image dataset. It also evaluates the efficiency of the network as it is influenced by different conditions. The efficiency is gauged by the error percentages accumulated by the network. Furthermore, statistical analysis is applied to the network in order to analyze the effectiveness.

The project showed that despite the adaptability of the neural network, it is still unable to remove the error completely. While neural networks are useful, they cannot be relied on completely. However, there is a tradeoff between error and flexibility in the network. While the error has the potential to be removed from the testing of the neural network, the network would be over fitted to the dataset it is being trained and tested on so that it would lose its ability to successfully analyze similar forms of data.

**Background Research**

*Neural Networks versus Conventional Approaches*

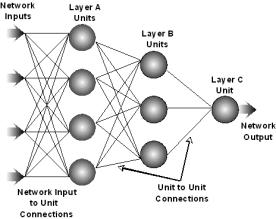
Neural networks take a different approach to problem solving than conventional algorithmic problem solving. Algorithmic problem solving requires a fixed set of actions to determine a solution, and if absent, an algorithmic function for such a problem is impossible, restricting the problem solving capabilities of conventional computing. Neural networks, in contrast, learn by example and cannot be programmed to perform a specific task, allowing a computer system to approximate an otherwise unknown function. Neural networks, as a result, are able to perform adaptive learning, self-organization, real time operation, as well as fault tolerance. The reliability of a neural network fails to match that of a conventional algorithm, as operation under certain conditions can be unpredictable. Often, to combat uncertainties in the neural network and the limitations of an algorithmic approach, the two problem solving methods are combing to perform a task at a high efficiency.

*Neural Network Similarity to Humans*

Neural networks are a form of artificial intelligence that derives from the structure of biological nervous systems and data processing methods within those systems. Both the brain and neural networks are composed of a large number of processing elements, referred to as neurons, which are interconnected by weights (in artificial neural networks) or axons and dendrites (in biological neural networks), which work together with the neurons to solve problems. Artificial networks and an organism's nervous system learn by example. These two structures update their weights or connections in response to inputs and whatever result is desired.Whereas artificial networks are typically data intensive and thus limited to several hundred units, biological neurons can consist of 10,000 individual inputs, immensely more complex. The complexity of the existing neural network is limited by the computing power of the computers or artificial systems in place today.

*Neural Networks*

To account for the lack of an algorithm, neural networks attempt to discern patterns from a dataset. Although any dataset is able to work, a larger dataset is typically required to allow a more adaptive nature for the neural network. The neural network takes in a set of inputs and passes on the values through a series of numbers termed “weights”, which then pass on an altered value to neurons in subsequent layers (either hidden or output). Once the network outputs values, the results are compared to the desired output, and the network's weights are adjusted accordingly, through the method of backpropagation. The feed forward neural network is the simplest type of an artificial neural network utilized. Information moves in a single direction, forward, from the input nodes, through hidden nodes (if any) and to output nodes. There are no cycles or loops in the network. Feed forward networks are the most popular and widely used function-modeling structures that reflect a dataset.



**Backpropagation**

Backpropagation, or backward propagation of error, is the most common algorithm for training neural networks. It is used to find a function that best models the input data. The goal of the backpropagation algorithm is to train any neural network such that it can learn to create any arbitrary map from inputs to outputs. From the error between the desired output and the actual output, the network readjusts its weights to reduce the error being produced. Backpropagation is most useful for feed forward networks which have no feedback.

**Nudging**

Nudging was implemented into the network. The point of nudging is to ensure that the network does not get stuck when training. At times because of the randomly distributed starting weights, the training of the network using backpropagation would get stuck in a local minimum that does not reach the desired error for convergence rate. If this occurs to the network, then nudging needs to occur in order to converge at the predetermined minimum error value. Reducing the error of the network is extremely important to implement it for diagnosis of breast cancer because of the life threatening issue of the disease.

**Haar Wavelet Transform**

Haar wavelet is a subset of a larger concept known as wavelet transform. Wavelet transformations are meant to change the time-frequency of an image. The transformation however only occurs in the time extension while maintaining the shape. A purpose of wavelet transformation is to compress images while reducing loss of data and is used over Fourier transformations because it is able to capture both frequency and location information. Inside wavelet transforms, there are two sub-categories, discrete and continuous. Discrete is represented by integers while continuous can be represented over an entire range of numbers. Discrete is preferable for image recognition because it is used to represent pixel values which are whole numbers. These pixel values once altered by wavelet transform are then used as input for a neural network.

**Implementation**

*Data processing*

The training dataset of the original Wisconsin Dataset was altered first by randomly excluding 68 data points, or 10% percent, of the 683 original dataset. 10 different neural networks were created by training them on the training dataset and their weights were then saved. These sets of weights are radically different from each other because of the existence of local minima, random weight space, and a preset convergence at .004% error.

The 68 data points excluded were used as the testing data points. These data points were independently ran through the networks and then the probability of malignancy was recorded. A step function was then implemented to heavily weight the results of the network towards the malignant output. If the output was greater than .05 then the network would automatically consider the output for that data point to be malignant.

Then, the output of the network was compared to the desired actual output. The average error for a single specific data point over all 10 networks next. Furthermore the average error for the total network is calculated by the average of the error for each data point. These calculations result in the total error of the network.

The entire equation is given by,



The same process occurred for the diagnostic Wisconsin Dataset, where 59 data points of also 10% were excluded for training. However, a difference that occurred is that the training would continue until the testing dataset reached an error of 2%, which is then considered the total error over all ten networks. The final or most optimal step function point was at 0.3. Because of the life threatening issue diagnosing breast cancer, the output is more weighted towards the malignant side not only to reduce error but to reduce the probability of a type 2 statistical error occurring. Further experiments occurred on this dataset to determine the best network variables, such as momentum and accelerative learning rate, for this network and specific dataset.

*Experiments (Wisconsin Breast Cancer Diagnostic Dataset)*

Control Experiment, Step Experiment, Accelerative Learning Rate Experiment, Momentum Experiment, Covariance Experiment, Conclusion Experiment

**Datasets**

*Breast Cancer Wisconsin (Original) Dataset – P-FNA Test*

This dataset is numerical and multivariate with integer attribute values. The creator was Dr. William H. Wolberg at the University of Wisconsin in Madison, Wisconsin. The diagnostic test created for this specific dataset is called a P-FNA test, proportional fine needle aspiration test.

The dataset used had 9 input attributes, each from a range of 1 to 10. There were a total of 699 data points. However, 16 of the points had inconsistencies where a question mark stood in place of a number. The 16 data points thus were excluded from both the training and the testing of the neural network. 10 percent of the data was used for testing while the other 90 percent were used for training. While the data had the output of 2 for benign and 4 for malignant, during the testing of the network these numbers were changed to 0 and 1 for benign and malignant respectively. This is because the logistic function can only output from a range of -1 to 1. For the actual dataset, 65.5% were benign and the other 35.5 percent were malignant. The source also claimed that there is also a 5% discrepancy in the dataset.

The input values are:

* Clump Thickness                (1 – 10)
* Uniformity of Cell Size       (1 – 10)
* Uniformity of Cell Shape    (1 – 10)
* Marginal Adhesion              (1 – 10)
* Single Epithelial Cell Size   (1 – 10)
* Bare Nuclei                    (1 – 10)
* Bland Chromatin                (1 – 10)
* Normal Nucleoli                (1 – 10)
* Mitoses                        (1 – 10)

*Breast Cancer Wisconsin (Diagnostic) Dataset – D-FNA Test*

The dataset used had 30 input attributes, each represented by a real value with four significant digits. Ten features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image. The mean, standard error, and largest of these features were computed for each image, resulting in 30 real valued features. There were a total of 569 data points. 10 percent of the data was used for testing while the other 90 percent were used for training. For the actual dataset, 62.7% were benign and the other 37.3% were malignant.

The input values are

* Radius - Mean of distances from center to points on the perimeter
* Texture - Standard deviation of gray-scale values
* Perimeter - Perimeter of the nucleus
* Area - Area of nucleus
* Smoothness - Local variation in radius lengths
* Compactness - Perimeter2 /Area - 1.0
* Concavity - Severity of concave portions of the contour
* Concave points - Number of concave portions of the contour
* Symmetry
* Fractal dimension – An objective measure of the complexity of the tissue of the biopsy

The mean, standard error, and largest of these features were computed for each image, resulting in 30 real valued features.

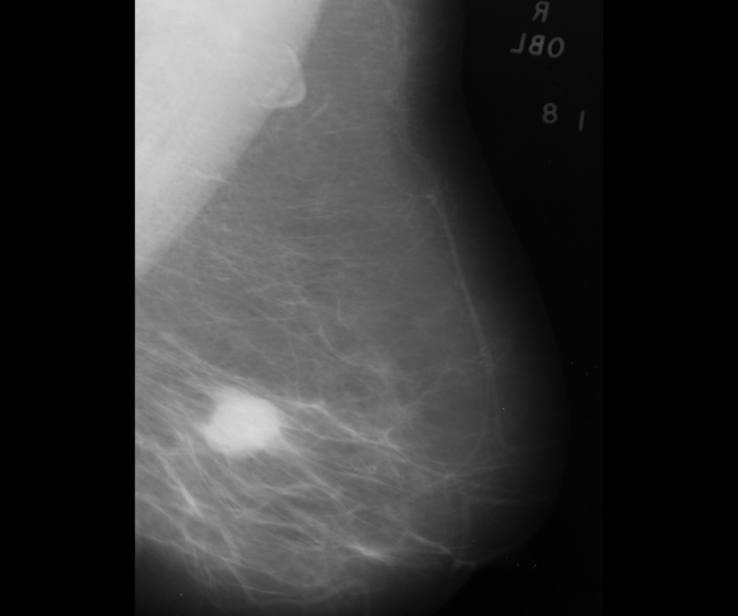
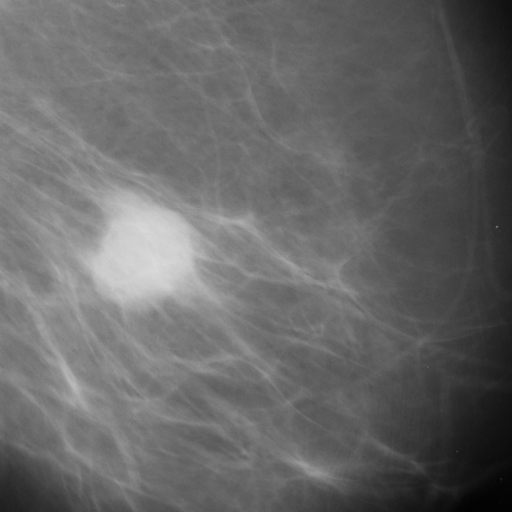
*Mammographic Image Analysis Society (MIAS) Database*

This dataset is composed of images from mammographies. The creator was the Mammographic Image Analysis Society. The diagnostic test created for this specific dataset is called a MID test, mammography image diagnostic test.

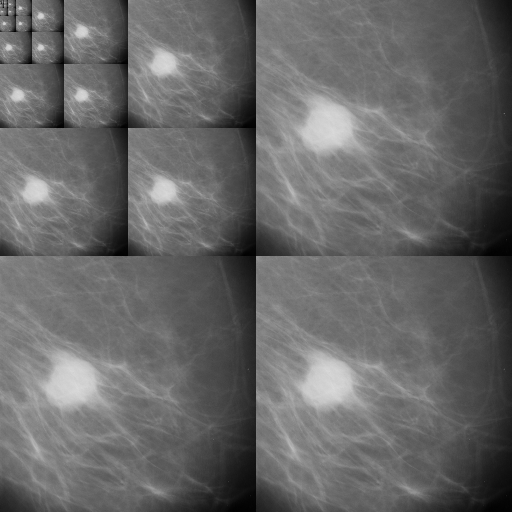
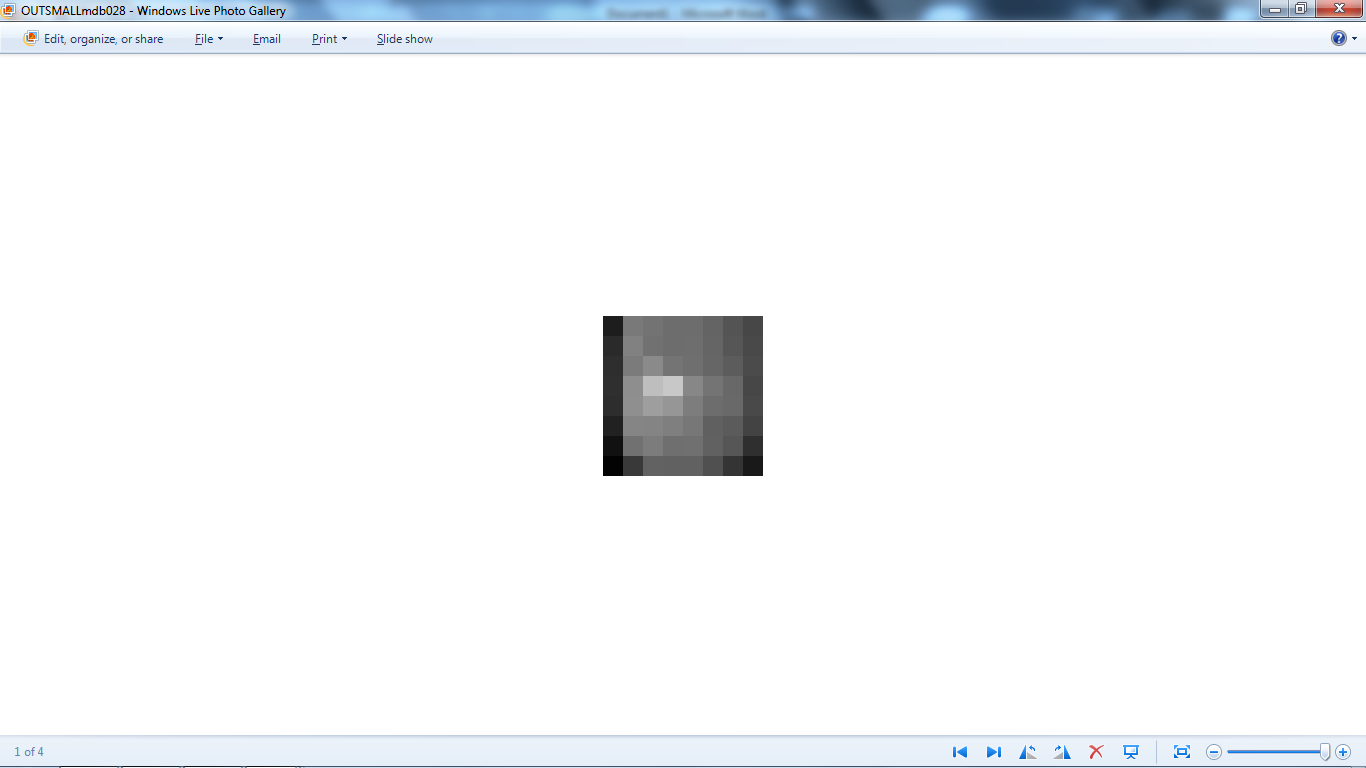
The dataset is composed of images with 200 micron pixel edges. Each image size is 1024x1024 pixels. The dataset used for input to the neural network is called a mini-MIAS database created by J. Suckling who reduced the resolution of the original MIAS database. There were a total of 322 data images. 18.2 percent of the data was used for testing while the other 81.8 percent were used for training. However, only a small portion of the actual images were used. The breasts that were in the dataset had three distinct background tissues: fatty, fatty-glandular, and dense-glandular. Only images with fatty background tissue were used during the experiment to reduce a variable for the network. Also only images of left breasts were used to reduce the amount of uncertainty in the network because of the existence of black space on either side of the breast. Once these specific images were separated from the original dataset, a total of 22 images remained. Of these 11 were tumorous and 11 were normal (more normal ones existed but in order to achieve consistency during training and testing only 11 were used).

**Haar Wavelet**

Original Image Area of Interest

Haar Wavelet Transform Input Image for the Neural Network

**App and Website**

Both an app and a website were created to make the diagnosis more readily available to the general populace. In order to make the diagnosis more applicable, it needs to be available to a large number of people. Creating both an app and website does so because of the amount of technology existing in society today. Just by going onto their phones or onto Web, people are able to access these useful tools. Simple numerical inputs or image inputs are easily entered into either the app or website, making both more usable by the public. In addition, both the app and the website include all three diagnostic tests, which have different inputs based off the three datasets the neural network was trained on.

**Conclusion**

A successful neural network was created with correctly implemented backpropagation. The final total error calculated for the neural network is 3.2% for the Wisconsin Original Breast Cancer dataset and 2% for the diagnostic dataset. The image recognition portion of the diagnostic tests was also successful, with an extremely low error of 2%. However, one must consider that the error in the original dataset was 5% indicating that the network was able to adapt to the error but not completely. The gradient method used to determine the values of the weights is not very accurate because it locates only local minimum instead of the global error minimum would be more accurate. However, the problem with finding global minimum is that the network might over fit the data, causing it to only recognize the training dataset and lose its adaptive nature to recognize other potential data points, rendering it useless to model complex functions.

Although establishing potential for inaccuracy, the step function is an extremely important part of the data processing to recognize potential outliers. By weighting the output significantly, the network also becomes more accurate. Because the field the network is being applied to allows no room for error, minimizing the amount of error received is more beneficial.

Finally, the 10 networks differed greatly, caused by the random initialized weight values. Even though the values are centered near zero for the range, they are still randomly placed according to the Gaussian distribution, causing a variance in convergence. This demonstrates the importance of weights on the network and how different random weights are able to lead to different solutions, indicating the self-structuring nature of a neural network.

The different variables that affect the network such as momentum, accelerative learning rate, and network size were demonstrated to have large effects on the error produced by the network. It was also demonstrated that these variables are very codependent on one another as indicated by the momentum, accelerative learning rate, and covariance experimental tests.

The network also has many potential uses. It is an extremely flexible artificial intelligence structure that was able to adapt to three distinct datasets. It is not only applicable to breast cancer but any dataset that is or isn’t able to be modeled by conventional method. The project allowed an application and website to utilize a generated neural network to quickly and accurately diagnose cases of potential breast cancer. These tools can not only be available and used by the general population, but also used in the medical field to alleviate the workload on medical professionals while they diagnose breast cancer.