APPENDIX 1

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Mr. G.GOPU Research scholar, Department of Electrical and Electronic Engineering, PSG College of Technology, Coimbatore has conducted a live trial to acquire electrogastrography signal in accordance with the Helsinki declaration to our In and Out patient using the Electrogastrogram hardware and software setup for the research work titled "ACQUISITION AND ANALYSIS OF ELECTROGASTROGRAM FOR CLASSIFICATION OF DIGESTIVE SYSTEM DISORDERS". This hospital has provided him the patient database as also the opportunity to test on our patients under medical supervision.

I would like to state that such testing has provided useful for the doctors of gastroenterology department and helped as a diagnostic tool. Although the project is at a pilot scale the utility value is seen and the hospital would be pleased to welcome further work on the project or enhancements thereof in the future.

I WISH HIM ALL SUCCESS

Dr. M. Manoj, MD, DM
Medical Gastroenterologist

MedIndia Hospitals
MedIndia Institute of Medical Specialties (MIMS)
(Digestive Diseases Training & Research Centre)

ENDOSCOPY
Endo Ultrasoundography
Advanced Biliary
Pancreatic Endotherapy
"C" ARM with DSA
Interventional Radiology
Argon Plasma Coagulator
Capacitive Endoscopy
Embolagation Lab
G I Manometry
24hr Gastric pH Recorder
Hypogastic Brain test
G I ICU
Integrated Surgical Back-up
Endoscopy Learning Centre
G I Birth Centre

Chairman, MD
CHIEF INTERVENTIONAL GASTROENTEROLOGY

Dr. T.S. Chandrasekar,
MD, F. R. C. E.

MD, DM

MEDICAL GASTROENTEROLOGY
Dr. M. Maragush, MD, DM.

SURGICAL GASTROENTEROLOGY
Dr. V. G. Venkatesh, M.S., F. R. C. E.
Dr. A. K. Ravindran, MS.
Dr. D. M. Ramgopal Rao, MS.

RADIOLOGY
Dr. S. Vivekanandam, MBBS, DM, MR.

ANAESTHESIOLOGY
Dr. A. N. N. R. R. R. R. R.
Dr. P. Arul, MB.

CARDIOLOGY
Dr. V. Palaniappan, MD, DM.

HEMATOLOGY
Dr. A. P. H. Chinnaraj, MD, DM.

EUROPEAN
Dr. A. Nandagopal, MB, BS.

PSYCHIATRY
Dr. N. S. Meher, MD, DM.

Medline India Hospitals
89, Weerakanda Road, Hampankat, Coimbatore - 641 009.
Phone: 042 2-2255465, 2250047 E-mail: telemedIndia@yahoo.com, drmedIndia@gmail.com
www.qiendohospitals.com

Hotline: 98 680 200 500

Paving Ways Saving Lives
TO WHOMSOEVER IT MAY CONCERN

I wish to certify that Mr. G. GOPU Research scholar, Department of Electrical and Electronics Engineering, PSG College of Technology, Coimbatore has conducted a live trial with the cooperation of this hospital in accordance with the Helsinki declaration, on our In and Out patients using the Electrogastrogram hardware and software setup for the research work titled "ACQUISITION AND ANALYSIS OF ELECTROGASTROGRAM FOR CLASSIFICATION OF DIGESTIVE SYSTEM DISORDERS". This hospital has provided him the patient database as also the opportunity to test on our patients under medical supervision.

I would like to state that such testing has provided useful inputs for the doctors of gastroenterology department and helped as a diagnostic tool. Although the project is at a pilot scale the utility value is seen and the hospital would be pleased to welcome further work on the project or enhancements thereof in the future.

I wish him all success in his endeavor.

Dr. J. Krishnaveni, M.D, D.N.B, D.M
Gastroenterologist
PSG Hospital, Coimbatore
Reg.No: 54419

Dr. J. Krishnaveni, M.D (Fut), D.M
Gastroenterologist
PSG Hospitals
Coimbatore
APPENDIX 2

STATISTICAL PARAMETERS

Arithmetic Mean

The arithmetic mean (or simply the mean) of a list of numbers is the sum of all samples of EGG of the subject divided by the number samples of EGG of the subject. Here, the mean of the signal represents the average value of the amplitudes of the signal. Equation (3.1) defines the arithmetic mean.

\[ M = \sum \frac{X}{N} \]

where,

- \( M \) : Mean value of the EGG signal
- \( X \) : Sample of EGG
- \( N \) : Total number EGG samples

Root Mean Square Value

In mathematics, the root mean square (abbreviated RMS or rms), also known as the quadratic mean, is a statistical measure of the magnitude of a varying quantity (EGG sample). It is especially useful when signals have both positive and negative values. RMS value of an EGG data (or a continuous-time waveform) is the square root of the arithmetic mean (average) of the squares of the EGG data (or the square of the function that defines the continuous waveform).
In the case of a set of ‘$N$’ values of EGG data $\{X_1, X_2, \ldots, X_n\}$, the RMS value is given by Equation below

$$X_{\text{rms}} = \sqrt{\frac{X_1^2 + X_2^2 + \ldots + X_n^2}{N}}$$

**Median**

A median is a number dividing the higher half of an EGG sample from the lower half of an EGG sample. The median of a finite list of numbers can be found by arranging all the observations from the lowest value to the highest value and picking the middle one. If there are even numbers of observations, one often takes the mean of the two middle values.

**Standard Deviation**

The standard deviation of EGG is a measure of the spread of its values. It is defined as the square root of the variance. It is the RMS deviation of the values from their arithmetic mean. For example, in the EGG data $\{2.75, 3.15\}$ cpm, the mean is 2.95 cpm and the standard deviation is 2. This may be written as $\{2.75, 3.15\} \approx 2.95 \pm 2$.

Standard deviation is the most common measure of statistical dispersion, measuring how widely spread the values in a data set is. If the data points are all close to the mean, then the standard deviation is close to zero. If many data points are far from the mean, then the standard deviation is far from zero. If all data values are equal, then the standard deviation is zero. As its name implies it gives in a standard form an indication of the possible deviations from the mean. The below Equation represents this parameter.
Where,

\[ \sigma : \text{ standard deviation of the EGG signal,} \]

\[ X : \text{ single samples of EGG,} \]

\[ X_i : \text{ a list of samples of EGG: } X_1, X_2, X_3 \text{ etc.,} \]

\[ \overline{X} : \text{ The arithmetic mean of all the numbers in the list,} \]

\[ N : \text{ Total number EGG samples.} \]

**Variance**

The variance of a random variable is a measure of its statistical dispersion, indicating how the possible values of EGG data are spread around the expected value of EGG data. The expected value of EGG data shows the location of the distribution, the variance indicates the scale of the values of EGG data. Equation below represents the variance.

\[ \sigma^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{N} \]

where,

\[ \sigma^2 : \text{ variance of the EGG signal} \]

**Kurtosis**
Kurtosis is a measure of the ‘peakedness’ of the probability distribution of a real valued random variable. Higher kurtosis means more of the variance due to infrequent extreme deviations, as opposed to frequent modestly-sized deviations. A high kurtosis distribution has a sharper ‘peak’ and fatter ‘tails’, while a low kurtosis distribution has a more rounded with wider ‘shoulders’. For univariate of EGG data Y1, Y2… YN, Equation below defines kurtosis.

\[
\text{Kurtosis} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^4}{(N-1)\sigma^4}
\]

Where,

- Y : is the mean of EGG data
- σ : is the standard deviation of EGG data
- N : is the number of data points

**Correlation**

Correlation is one of the most common and most useful statistics. A correlation is a single number that describes the degree of relationship between two variables. Cross-correlation is a statistical measure timing the movements and proximity of alignment between two different information sets of a series of information or is a measure of similarity of two waveforms as a function of a time-lag applied to one of them. It is commonly used for searching a long-duration signal for a shorter, known feature. It also has applications in biosignal analysis, pattern recognition, neurophysiology, etc.

For continuous functions, f and g, the cross-correlation is defined in Equation as below
\[(f \ast g)(t) \text{def} \int_{-\infty}^{\infty} f^*(\tau)g(t + \tau)d\tau\]

Where,

\(f^*\): the complex conjugate of \(f\).

Similarly, for discrete functions, the cross-correlation is defined in Equation below

\[
CC = \frac{1}{N} \sum_{n} \sum_{m} x(n) y(n - m)
\]

Where,

- \(CC\) : Cross Correlation
- \(x(n)\) : Sample of normal EGG
- \(y(n)\) : Samples of disorders EGG
- \(N\) : Total number of samples

Cross-correlation is similar in nature to the convolution of two functions.
APPENDIX 3

RECORDING OF EGG SIGNALS WITH ITS STATISTICAL PARAMETERS FOR DIFFERENT EGG SUBJECTS

Figure A3.1(a) Preprandial Recording of EGG - Brady gastria
Figure A3.1 (b) Postprandial Recording of EGG - Bradygastria

Figure A3.2 (a) Preprandial Recording of EGG – Dyspepsia
Figure A3.2 (b) Postprandial Recording of EGG - Dyspepsia

Figure A3.3 (a) Preprandial Recording of EGG - Nausea
Figure A3.3 (b) Postprandial Recording of EGG - Nausea

Figure A3.4 (a) Preprandial Recording of EGG - Tachygastria
Figure A3.4(b) Postprandial Recording of EGG - Tachygastria

Figure A3.5 (a) Preprandial Recording of EGG - Ulcer
Figure A3.5 (b) Postprandial Recording of EGG – Ulcer

Figure A3.6 (a) Preprandial Recording of EGG - Vomiting
### Figure A3.6 (b) Postprandial Recording of EGG - Vomiting

### Table A3.1 Parameters of Bradygastria Subjects

<table>
<thead>
<tr>
<th>Sl.No.</th>
<th>Parameters</th>
<th>Preprandial Condition</th>
<th>Postprandial Condition</th>
<th>% of Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Mean</td>
<td>0.0215</td>
<td>0.0164</td>
<td>-31.1</td>
</tr>
<tr>
<td>2.</td>
<td>RMS</td>
<td>0.4504</td>
<td>0.3562</td>
<td>-26.4</td>
</tr>
<tr>
<td>3.</td>
<td>Median</td>
<td>0.062</td>
<td>0.0754</td>
<td>17.77</td>
</tr>
<tr>
<td>4.</td>
<td>Standard Deviation</td>
<td>0.4537</td>
<td>0.3588</td>
<td>-26.4</td>
</tr>
<tr>
<td>5.</td>
<td>Variance</td>
<td>0.2058</td>
<td>0.1287</td>
<td>-59.9</td>
</tr>
<tr>
<td>6.</td>
<td>Kurtosis</td>
<td>1.4003</td>
<td>1.5901</td>
<td>11.94</td>
</tr>
<tr>
<td>7.</td>
<td>Amplitude (V)</td>
<td>0.6970</td>
<td>0.5060</td>
<td>-37.7</td>
</tr>
<tr>
<td>8.</td>
<td>Frequency (Hz)</td>
<td>0.0290</td>
<td>0.0220</td>
<td>-31.8</td>
</tr>
<tr>
<td>9.</td>
<td>Power Spectrum (dB)</td>
<td>10.876</td>
<td>5.2957</td>
<td>-105</td>
</tr>
</tbody>
</table>

**Statistical Parameters**

- Arithmetic Mean
- RMS
- Standard Deviation
- Variance
- Median
- Kurtosis
- Mode
- Power Spectrum

**EGG signal characteristics**

- Amplitude (V)
- Frequency (Hz)
- Power Spectrum (dB)
Table A3.2 Parameters of Dyspepsia Subjects

<table>
<thead>
<tr>
<th>Sl.No.</th>
<th>Parameters</th>
<th>Preprandial Condition</th>
<th>Postprandial Condition</th>
<th>% of Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Mean</td>
<td>0.0218</td>
<td>0.0189</td>
<td>-15.3</td>
</tr>
<tr>
<td>2.</td>
<td>RMS</td>
<td>0.1877</td>
<td>0.3262</td>
<td>42.46</td>
</tr>
<tr>
<td>3.</td>
<td>Median</td>
<td>0.0279</td>
<td>0.0586</td>
<td>52.39</td>
</tr>
<tr>
<td>4.</td>
<td>Standard Deviation</td>
<td>0.1880</td>
<td>0.3284</td>
<td>42.75</td>
</tr>
<tr>
<td>5.</td>
<td>Variance</td>
<td>0.0353</td>
<td>0.1079</td>
<td>67.28</td>
</tr>
<tr>
<td>6.</td>
<td>Kurtosis</td>
<td>1.5384</td>
<td>1.4779</td>
<td>-4.09</td>
</tr>
</tbody>
</table>

EGG signal characteristics

<table>
<thead>
<tr>
<th></th>
<th>Preprandial Condition</th>
<th>Postprandial Condition</th>
<th>% of Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. Amplitude (V)</td>
<td>0.7970</td>
<td>0.5630</td>
<td>-41.6</td>
</tr>
<tr>
<td>8. Frequency (Hz)</td>
<td>0.0761</td>
<td>0.0359</td>
<td>-112</td>
</tr>
<tr>
<td>9. Power Spectrum (dB)</td>
<td>8.8521</td>
<td>3.1493</td>
<td>-181</td>
</tr>
</tbody>
</table>

Table A3.3 Parameters of Nausea Subjects

<table>
<thead>
<tr>
<th>Sl.No.</th>
<th>Parameters</th>
<th>Preprandial Condition</th>
<th>Postprandial Condition</th>
<th>% of Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Mean</td>
<td>0.0375</td>
<td>0.0275</td>
<td>-36.4</td>
</tr>
<tr>
<td>2.</td>
<td>RMS</td>
<td>0.1240</td>
<td>0.2744</td>
<td>54.81</td>
</tr>
<tr>
<td>3.</td>
<td>Median</td>
<td>0.0440</td>
<td>0.0793</td>
<td>44.51</td>
</tr>
<tr>
<td>4.</td>
<td>Standard Deviation</td>
<td>0.1192</td>
<td>0.2753</td>
<td>56.7</td>
</tr>
<tr>
<td>5.</td>
<td>Variance</td>
<td>0.0142</td>
<td>0.0758</td>
<td>81.27</td>
</tr>
<tr>
<td>6.</td>
<td>Kurtosis</td>
<td>1.8579</td>
<td>1.6207</td>
<td>-14.6</td>
</tr>
</tbody>
</table>

EGG signal characteristics

<table>
<thead>
<tr>
<th></th>
<th>Preprandial Condition</th>
<th>Postprandial Condition</th>
<th>% of Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. Amplitude (V)</td>
<td>0.5350</td>
<td>0.5280</td>
<td>-1.33</td>
</tr>
<tr>
<td>8. Frequency (Hz)</td>
<td>0.0760</td>
<td>0.0419</td>
<td>-81.4</td>
</tr>
<tr>
<td>9. Power Spectrum (dB)</td>
<td>6.7680</td>
<td>2.6887</td>
<td>-152</td>
</tr>
</tbody>
</table>
Table A3.4 Parameters of Tachygastria Subjects

<table>
<thead>
<tr>
<th>Sl.No.</th>
<th>Parameters</th>
<th>Preprandial Condition</th>
<th>Postprandial Condition</th>
<th>% of Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistical Parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Mean</td>
<td>0.0222</td>
<td>0.0154</td>
<td>-44.2</td>
</tr>
<tr>
<td>2.</td>
<td>RMS</td>
<td>0.0544</td>
<td>0.1176</td>
<td>53.74</td>
</tr>
<tr>
<td>3.</td>
<td>Median</td>
<td>0.0198</td>
<td>0.0132</td>
<td>-50</td>
</tr>
<tr>
<td>4.</td>
<td>Standard Deviation</td>
<td>0.0500</td>
<td>0.1176</td>
<td>57.48</td>
</tr>
<tr>
<td>5.</td>
<td>Variance</td>
<td>0.0025</td>
<td>0.0138</td>
<td>81.88</td>
</tr>
<tr>
<td>6.</td>
<td>Kurtosis</td>
<td>2.2328</td>
<td>1.9426</td>
<td>-14.9</td>
</tr>
<tr>
<td></td>
<td>EGG signal characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Amplitude (V)</td>
<td>0.6580</td>
<td>0.5810</td>
<td>-13.3</td>
</tr>
<tr>
<td>8.</td>
<td>Frequency (Hz)</td>
<td>0.1120</td>
<td>0.0827</td>
<td>-35.4</td>
</tr>
<tr>
<td>9.</td>
<td>Power Spectrum (dB)</td>
<td>14.836</td>
<td>8.6260</td>
<td>-72</td>
</tr>
</tbody>
</table>

Table A3.5 Parameters of Ulcer Subjects

<table>
<thead>
<tr>
<th>Sl.No.</th>
<th>Parameters</th>
<th>Preprandial Condition</th>
<th>Postprandial Condition</th>
<th>% of Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistical Parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Mean</td>
<td>0.0323</td>
<td>0.0220</td>
<td>-46.8</td>
</tr>
<tr>
<td>2.</td>
<td>RMS</td>
<td>0.0596</td>
<td>0.2833</td>
<td>78.96</td>
</tr>
<tr>
<td>3.</td>
<td>Median</td>
<td>0.0292</td>
<td>0.0390</td>
<td>25.13</td>
</tr>
<tr>
<td>4.</td>
<td>Standard Deviation</td>
<td>0.0504</td>
<td>0.2848</td>
<td>82.3</td>
</tr>
<tr>
<td>5.</td>
<td>Variance</td>
<td>0.0025</td>
<td>0.0811</td>
<td>96.92</td>
</tr>
<tr>
<td>6.</td>
<td>Kurtosis</td>
<td>2.3947</td>
<td>1.4666</td>
<td>-63.3</td>
</tr>
<tr>
<td></td>
<td>EGG signal characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Amplitude (V)</td>
<td>0.6490</td>
<td>0.4610</td>
<td>-40.8</td>
</tr>
<tr>
<td>8.</td>
<td>Frequency (Hz)</td>
<td>0.0567</td>
<td>0.0340</td>
<td>-66.8</td>
</tr>
</tbody>
</table>
Table A3.6  Parameters of Vomiting Subjects

<table>
<thead>
<tr>
<th>Sl.No.</th>
<th>Parameters</th>
<th>Preprandial Condition</th>
<th>Postprandial Condition</th>
<th>% of Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistical Parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Mean</td>
<td>0.0424</td>
<td>0.0245</td>
<td>-73.1</td>
</tr>
<tr>
<td>2.</td>
<td>RMS</td>
<td>0.0819</td>
<td>0.2262</td>
<td>63.79</td>
</tr>
<tr>
<td>3.</td>
<td>Median</td>
<td>0.0429</td>
<td>0.0302</td>
<td>-42.1</td>
</tr>
<tr>
<td>4.</td>
<td>Standard Deviation</td>
<td>0.0707</td>
<td>0.2267</td>
<td>68.81</td>
</tr>
<tr>
<td>5.</td>
<td>Variance</td>
<td>0.005</td>
<td>0.0514</td>
<td>90.27</td>
</tr>
<tr>
<td>6.</td>
<td>Kurtosis</td>
<td>0.0424</td>
<td>0.0245</td>
<td>-73.1</td>
</tr>
<tr>
<td></td>
<td>EGG signal characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Amplitude (V)</td>
<td>0.5890</td>
<td>0.4560</td>
<td>-29.2</td>
</tr>
<tr>
<td>8.</td>
<td>Frequency (Hz)</td>
<td>0.0993</td>
<td>0.0469</td>
<td>-112</td>
</tr>
</tbody>
</table>
APPENDIX 4

TRAINING ALGORITHM

BP-MRAN is trained using nine different algorithms namely,

- Gradient Descent with momentum and adaptive learning rate Backpropagation (GDX)
- Resilient Backpropagation (RP)
- Conjugate Gradient Backpropagation with Fletcher-Reeves updates (CGF)
- Conjugate gradient Backpropagation with Polak-Ribiere updates (CGP)
- Conjugate gradient Backpropagation with Powell-Beale restarts (CGB)
- Scaled conjugate gradient Backpropagation (SCG)
- BFGS quasi-Newton Backpropagation (BFG)
- One Step Secant Backpropagation (OSS)
- Levenberg-Marquardt Backpropagation (LM)

In all the training algorithms of BPNN given below the following parameters are applied as inputs i.e.

Net : Neural network.
tr : Initial training record created by TRAIN.
trainV : Training data created by TRAIN.
valV : Validation data created by TRAIN.
testV : Test data created by TRAIN.

and in return it gives,

net : Trained network.

tr : Training record of various values over each epoch.

The training of the above mentioned algorithms stops when the maximum number of epochs (repetitions) is reached.

A brief description of various algorithms used in BPNN for training is presented below.

**Variable Learning Rate BP with Momentum (traingdx)**

The learning rate parameter is used to determine how fast the BP method converges to the minimum solution. The larger the learning rate, the bigger the step and the faster the convergence. However, if the learning rate is made too large the algorithm will become unstable. On the other hand, if the learning rate is set to too small, the algorithm will take a long time to converge. To speed up the convergence time, the variable learning rate gradient descent BP utilizes larger learning rate $\alpha$ when the neural network model is far from the solution and smaller learning rate $\alpha$ when the neural net is near the solution. The new weight vector $w_{k+1}$ is adjusted the same as in the gradient descent with momentum above but with a varying $\alpha_k$. Typically, the new weight vector $w_{k+1}$ is given by equations below

$$w_{k+1} = w_k - \alpha_{k+1}g_k + \mu w_{k-1}$$

$$\alpha_{k+1} = \beta \alpha_k$$
where,

\[ \beta : \begin{cases} 0.7, & \text{if the new error is greater than 1.04 (old error)} \\ 1.05, & \text{if the new error is lesser than 1.04 (old error)} \end{cases} \]

\[ \mu : \text{Momentum factor} \]

\[ \alpha_k : \text{Learning rate} \]

\[ g_k : \text{Current gradient} \]

\[ w_k : \text{Vector of current weights and biases} \]

**Syntax:**

\[ [\text{net, tr}] = \text{traingdx (net, tr, trainV, valV, testV)} \]

\[ \text{info} = \text{traingdx ('info')} \]

**Resilient Backpropagation (trainrp)**

The purpose of the resilient backpropagation training algorithm is to eliminate the harmful effects of the magnitudes of the partial derivatives caused when using steepest descent to train a multilayer network with sigmoid functions, since the gradient can have a very small magnitude; and therefore, cause small changes in the weights and biases, even though the weights and biases are far from their optimal values. Only the sign of the derivative is used to determine the direction of the weight update; the magnitude of the derivative has no effect on the weight update. The size of the weight change is determined by a separate update value.

The update value for each weight and bias is increased by a factor \( \text{delt_inc} \) (Increment to weight change) whenever the derivative of the performance function with respect to that weight has the same sign for two
successive iterations. The update value is decreased by a factor delt_dec (Decrement to weight change) whenever the derivative with respect that weight changes sign from the previous iteration. If the derivative is zero, then the update value remains the same. Whenever the weights are oscillating the weight change will be reduced. If the weight continues to change in the same direction for several iterations, then the magnitude of the weight change will be increased.

The performance of RP algorithm is not very sensitive to the settings of the training parameters. It is generally much faster than the standard steepest descent algorithm. It also has the nice property that it requires only a modest increase in memory requirements. It is need to store the update values for each weight and bias, which is equivalent to storage of the gradient.

Syntax:

```
[net, tr, Ac, El] = trainrp (net, tr, trainV, valV, testV)
```

```
info = trainrp('info')
```

**Conjugate Gradient BP**

The basic BP algorithm adjusts the weights in the steepest descent direction. This is the direction in which the performance function is decreasing most rapidly. Although the function decreases most rapidly along the negative of the gradient, this does not necessarily produce the fastest convergence. In the conjugate gradient algorithms, a search is performed along conjugate directions, which produces generally faster convergence than steepest descent directions. In the conjugate gradient algorithms, the step size is adjusted at each iteration. A search is made along the conjugate gradient direction to determine the step size which will minimize the performance
function along that line. The different search function related to Conjugate Gradient included in the toolbox as follows

a. Fletcher-Reeves Update (traincgf)

b. Polak-Ribiére Update (traincgp)

c. Powell-Beale Restarts (traincgb)

d. Scaled Conjugate Gradient (trainscg)

All of the conjugate gradient algorithms start out by searching in the steepest descent direction (negative of the gradient) on the first iteration as given equation below \( P_0 = -g_0 \)

where,

\( P_0 \): Initial search gradient

\( g_0 \): Initial Gradient

A line search is then performed to determine the optimal distance to move along the current search direction as given in the equation below

\[ w_{k+1} = w_k + \alpha_k p_k \]

Then the next search direction is determined so that it is conjugate to previous search directions. The general procedure for determining the new search direction is to combine the new steepest descent direction with the previous search direction as given by equation below

\[ p_k = -g_k + \beta_k p_{k-1} \]
where,

\[ \beta_k : \text{Constant} \]

\[ w_k : \text{Current weight vector} \]

\[ w_{k+1} : \text{Next weight vector} \]

\[ \alpha_k : \text{Learning rate} \]

\[ p_k : \text{Current search direction} \]

\[ p_{k-1} : \text{Previous search direction} \]

The various versions of conjugate gradient are distinguished by the manner in which the constant \( \beta_k \) is computed.

**Fletcher-Reeves Update (traincgf)**

The Fletcher-Reeves update procedure is given in the equation below

\[
\beta_k = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}
\]

This is the ratio of the norm squared of the current gradient to the norm squared of the previous gradient.

**Syntax:**

\[
\text{[net, tr]} = \text{traincgf}(	ext{net, tr, trainV, valV, testV})
\]

\[
\text{info} = \text{traincgf}('\text{info}')
\]
Polak-Ribiére Update (traincgp)

Another version of the conjugate gradient algorithm was proposed by Polak and Ribiere. As with the Fletcher-Reeves algorithm, the search direction at each iteration is determined by equation below

\[ p_k = -g_k + \beta_k p_{k-1} \]

For the Polak-Ribiére update, the constant \( \beta_k \) is computed by the equation below

\[ \beta_k = \frac{\Delta g_k^T g_k}{g_{k-1}^T g_{k-1}} \]

This is the inner product of the previous change in the gradient with the current gradient divided by the norm squared of the previous gradient.

Syntax:

\[ [\text{net, tr}] = \text{traincgp}(\text{net, tr, trainV, valV, testV}) \]

\[ \text{info} = \text{traincgp}('\text{info}') \]

Powell-Beale Restarts (traincgb)

In all conjugate gradient algorithms, the search direction will be periodically reset to the negative of the gradient. The standard reset point occurs when the number of iterations is equal to the number of network Parameters (weights and biases), but there are other reset methods that can improve the efficiency of training. One such reset method was proposed by Powell based on an earlier version proposed by Beale. This technique will restart if there is very little orthogonality left between the current gradient and
the previous gradient. This is tested with the following inequality given in equation below.

\[ g_k^T g_k \geq 0.2 \| g_k \|^2 \]

where,

\[ g_k \] : Current gradient

If this condition is satisfied, the search direction is reset to the negative of the gradient.

**Syntax:**

\[
[\text{net}, \text{tr}] = \text{traincgp}(\text{net}, \text{tr}, \text{trainV}, \text{valV}, \text{testV})
\]

\[
\text{info} = \text{traincgp}('\text{info}')
\]

**Scaled Conjugate Gradient (trainscg)**

Each of the conjugate gradient algorithms that have been discussed so far requires a line search at each iteration. This line search is computationally expensive, since it requires that the network response to all training inputs be computed several times for each search. The scaled conjugate gradient algorithm (SCG), developed by Moller was designed to avoid the time-consuming line search. This algorithm is too complex to explain in a few lines, but the basic idea is to combine the model-trust region with the conjugate gradient approach.

The trainscg routine may require more iteration to converge than the other conjugate gradient algorithms, but the number of computations in each iteration is significantly reduced because no line search is performed. The storage requirements for the scaled conjugate gradient algorithm are about the same as those of Fletcher-Reeves.
Syntax:

\[
[\text{net}, \text{tr}, \text{Ac}, \text{El}] = \text{trainscg}([\text{net}, \text{tr}, \text{trainV}, \text{valV}, \text{testV}])
\]

\[
\text{info} = \text{trainscg}('\text{info}')
\]

**BFGS Quasi-Newton Algorithms (trainbfg)**

Newton’s method is an alternative to the conjugate gradient methods for fast optimization. Newton’s method often converges faster than conjugate gradient methods. The weight update for the Newton’s method is given in the equation below

\[
w_{k+1} = w_k - A_k^{-1} g_k
\]

where,

\[A_k\] : the Hessian matrix of the performance index at the current values of the weights and biases

If \[A_k\] is large, it is complex and time consuming to compute \(w_{k+1}\). There is a class of algorithms based on the works of Broyden, Fletcher, Goldfarb and Shanno (BFGS) that are based on Newton’s method but which do not require intensive calculation. This new class of method is called quasi-Newton method. The new weight \(w_{k+1}\) is computed as a function of the gradient and the current weight \(w_k\).

Syntax:

\[
[\text{net}, \text{TR}] = \text{trainbfg}([\text{net}, \text{TR}, \text{trainV}, \text{valV}, \text{testV}])
\]

\[
\text{info} = \text{trainbfg}('\text{info}')
\]
**One Step Secant Algorithm (trainoss)**

Since the BFGS algorithm requires more storage and computation in each iteration than the conjugate gradient algorithms, there is need for a secant approximation with smaller storage and computation requirements. The one step secant (OSS) method is an attempt to bridge the gap between the conjugate gradient algorithms and the quasi-Newton (secant) algorithms. This algorithm does not store the complete Hessian matrix; it assumes that at each iteration, the previous Hessian was the identity matrix. This has the additional advantage that the new search direction can be calculated without computing a matrix inverse.

This algorithm requires less storage and computation per epoch than the BFGS algorithm. It requires slightly more storage and computation per epoch than the conjugate gradient algorithms. It can be considered a compromise between full quasi-Newton algorithms and conjugate gradient algorithms.

**Syntax:**

\[
[\text{net}, \text{tr}, \text{Ac}, \text{El}] = \text{trainoss} \left(\text{net}, \text{tr}, \text{trainV}, \text{valV}, \text{testV}\right)
\]

\[
\text{info} = \text{trainoss} \left(\text{info}\right)
\]

**Levenberg-Marquardt (trainlm)**

The Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares (as is typical in training feedforward networks), then the Hessian matrix can be approximated as in equation below

\[
H = J^TJ
\]
and the gradient can be computed as in equation below

\[ g = J^T e \]

where,

\( J \) : Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases

\( e \) : Vector of network errors

The Jacobian matrix can be computed through a standard backpropagation technique. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update as in Equation below

\[ x_{k+1} = x_k - \left[ J^T J + \mu I \right]^{-1} J^T e \]

When the scalar \( \mu \) is zero, this is similar to Newton’s method, using the approximate Hessian matrix. When \( \mu \) is large, this becomes gradient descent with a small step size. Newton’s method is faster and more accurate near an error minimum, so the aim is to shift towards Newton’s method as quickly as possible. Thus, \( \mu \) is decreased after each successful step (reduction in performance function) and is increased only when a tentative step would increase the performance function. In this way, the performance function will always be reduced at each iteration of the algorithm. This algorithm appears to be the fastest method for training moderate-sized feedforward neural networks (up to several hundred weights). It also has a very efficient MATLAB implementation, since the solution of the matrix equation is a built-in function.

**Syntax:**

```matlab
[net, tr] = trainlm (net, tr, trainV, valV, testV)
```

```
info = trainlm ('info')
```