Methodology Article

Breathomics for Gastric Cancer Classification Using Back-propagation Neural Network

D. Arul Pon Daniel, K. Thangavel[1]

Department of Computer Science, Loyola College, Mettala, Namakkal, [1]Department of Computer Science, Periyar University, Salem, Tamil Nadu, India

ABSTRACT

Breathomics is the metabolomics study of exhaled air. It is a powerful emerging metabolomics research field that mainly focuses on health-related volatile organic compounds (VOCs). Since the quantity of these compounds varies with health status, breathomics assures to deliver noninvasive diagnostic tools. Thus, the main aim of breathomics is to discover patterns of VOCs related to abnormal metabolic processes occurring in the human body. Classification systems, however, are not designed for cost-sensitive classification domains. Therefore, they do not work on the gastric carcinoma (GC) domain where the benefit of correct classification of early stages is more than that of later stages, and also the cost of wrong classification is different for all pairs of predicted and actual classes. The aim of this work is to demonstrate the basic principles for the breathomics to classify the GC, for that the determination of VOCs such as acetone, carbon disulfide, 2-propanol, ethyl alcohol, and ethyl acetate in exhaled air and stomach tissue emission for the detection of GC has been analyzed. The breath of 49 GC and 30 gastric ulcer patients were collected for the study to distinguish the normal, suspected, and positive cases using back-propagation neural network (BPN) and produced the accuracy of 93%, sensitivity of 94.38%, and specificity of 89.93%. This study carries out the comparative study of the result obtained by the single- and multi-layer cascade-forward and feed-forward BPN with different activation functions. From this study, the multilayer cascade-forward outperforms the classification of GC from normal and benign cases.

Key words: *Breath Analysis, Human Body, Metabolomics, Neural Networks, Sensitivity and Specificity, Stomach Cancer, Stomach Ulcer, Volatile Organic Compounds*

INTRODUCTION

Recently, analytical methods for measuring volatile organic compounds (VOCs) in exhaled air with high resolution and high throughput have been extensively developed. Yet, the application of machine learning methods for fingerprinting VOC profiles in the breathomics is still in its infancy. A recent literature suggests that the potential utility of breath analysis is an alternative noninvasive methodology.

In clinical medicine, reaching a conclusion about a patient's symptoms, when presented with complex and sometimes contradictory clinical information, is really difficult. A clinician usually makes decisions based on a set of measurements and observations about a patient and evaluates all the factors subjectively to reach a diagnosis. However, it is obvious that clinicians may have great difficulty in analyzing enormous amount of clinical and histopathological data. Therefore, more sophisticated quantitative techniques are needed to help clinicians consider all the data and make better diagnoses. Some sophisticated quantitative techniques are proposed to the doctors by computer scientists through machine-learning techniques to help in this decision-making process.[1]

Breath diagnostics, the measurement of volatile chemicals in human breath, is currently receiving attention as a technique for the detection of disease which, being noninvasive in nature, is particularly suited to screening for presymptomatic disease in healthy populations.[2] An entirely noninvasive methodology, breath analysis, has the potential to deliver accurate and reproducible diagnostic tests without risk to the patient, making it ideal for population-based health screening as well as individual testing in response to symptom occurrence. Breath analysis

Address for correspondence:
Prof. D. Arul Pon Daniel, Department of Computer Science,
Loyola College, Mettala, Namakkal, Tamil Nadu, India.
E-mail: apdaniel86@yahoo.com

relies on the fact that disease states alter cellular metabolite levels, these being transferred to the bloodstream and, for volatile compounds, subsequently discerned in the breath.[3]

Gastric carcinoma (GC) is the second most common cause of cancer-related deaths among Indian men and women.[4] GC ranks among the five most common cancer among young Indian men and women (aged 15–44 years) based on a study from Karnataka.[5] It has been estimated that the number of new GC cases is about 34,000 (with a male predominance ratio of 1:2) with a progressive increase postulated such that by the year 2020, there would be approximately 50,000 new GC cases annually in India.[6]

The Gastric Carcinoma Domain

GC is a disease in which cancer (malignant) cells are found in the tissues of the stomach. The stomach is a J-shaped organ in the upper abdomen where the food is digested. Food reaches the stomach through a tube called the esophagus that connects the mouth to the stomach. After leaving the stomach, partially digested food passes into the duodenum then the small intestine and then into the large intestine called the colon. Sometimes, cancer can be in the stomach for a long time and can grow very large before it causes any symptom. In the early stages of the stomach cancer, a patient may have indigestion and stomach discomfort, a bloated feeling after eating, mild nausea, loss of appetite, or heartburn. In more advanced stages of cancer of the stomach, the patient may have blood in stool and vomiting, weight loss, or pain in the stomach. Some factors that increase the chances of getting stomach cancer are a stomach disorder, called atrophic gastritis, disorder of the blood, called anemia, or a hereditary condition of growths, called polyps, in the large intestine. Stomach cancer is difficult to detect in its early stages because its early symptoms are absent or mild. Unfortunately, this is a highly aggressive cancer and the overall survival rate is very low. The chance of recovery (prognosis) and the choice of treatment depend on the stage of cancer, whether it is just in the stomach or it has spread to other places, and the patient's general state of health.[1]

Classification of Gastric Carcinoma

If there are symptoms of cancer, a physician will usually order an upper gastrointestinal X-ray or he may also look inside the stomach with a thin, lighted tube called a gastroscope. This procedure is called gastroscopy, and it is useful in the detection of most stomach cancer. For this test, the gastroscope is inserted through the mouth and guided into the stomach and the stomach mucosa is examined. According to the Gastroenterological Endoscopy Society, based on the visual inspection of the mucosal surface of the patient's stomach, GC is classified mainly into three categories: early GC (EGC) and advanced GC (AGC)

and the remaining ones which cannot be included in these categories.[7]

EGC is defined as GC confined to the mucosa or submucosa, regardless of the presence or absence of lymph node.[8] On the other hand, in AGC, as defined by Bormann, the tumor is invaded into the proper muscle layer beyond the stomach.[9] Moreover, knowledge of these types permits a preliminary assessment of tumor spread in stomach.

Chemical analysis of the breath samples showed that five VOCs (2-propenenitrile, 2-butoxyethanol, furfural and 6-methyl-5-hepten-2-one and isoprene) were significantly elevated in patients with GC and/or peptic ulcer as compared with less severe gastric conditions. The encouraging preliminary results presented here have initiated a multicenter clinical trial with considerably increased sample size to confirm the observed breath prints.

This study has been organized into five sections. Section 2 presents the experimental setup of Breath analysis. Section 3 elaborates the methodology of collecting Breath samples from the volunteers. Section 4 discusses the experimental method and materials. Section 5 presents the experimental analysis results. In Section 6, conclusion and further research scope are presented.

EXPERIMENTAL SETUP

Study Design

The primary aim of this study was to distinguish GC patients from patients with benign gastric conditions who may present similar clinical symptoms. The secondary aim was to distinguish subpopulations in the malignant and nonmalignant study groups. This study with a limited patient group of 161 (out of 236 patients after application of the exclusion criteria) was designed as a feasibility test of a nonmaterial-based breath test for GC, with a more realistic ration of malignant to nonmalignant gastric conditions.

Sensor Array

Three screen-printed commercially available metal oxide semiconductor gas sensor arrays are used to construct the proposed Breath analyzer. The gas sensors are manufactured and commercialized by Figaro USA Inc. The resulting array, populated by sensor devices tagged by the manufacturer as TGS813, TGS822, and TGS2620, is placed into a test chamber. The obtained sensor element is mounted onto a stainless steel substrate with head of chlorinated polyvinyl chloride, and then connected by lead wires to the pins of the sensor package. To generate the required dataset, connect the said test chamber to a data acquisition card (DAQ), which provides versatility for conveying the chemical compounds of interest at the desired concentrations to the sensing chamber. The

response of the gas sensor array was measured when the operating temperature of sensors which, according to the deterministic one-to-one look-up table provided by the manufacturer (Figaro USA Inc., http://www.figarosensor.com), is attained through a built-in heater that is driven by an external DC voltage source set at 5V. The sensor response is read-out in the form of the resistance across the active layer of each sensor; hence, each measurement produces a 6-channel time series sequence. The DAQ collects data from the gas sensors and controls the analog voltage signal to every sensor heater. The experimental setup is shown in Figure 1.

COLLECTION OF BREATH SAMPLES

Exhaled alveolar breath was collected in a controlled manner, none of the volunteers consumed food, tobacco, or alcohol during an (overnight) 12 h interval before the breath collection. All volunteers were asked to rest for 1 h before the breath sampling and did not perform heavy physical exercise 24 h before taking the breath sample. All breath samples were collected in the same clinical environment and in duplicates (for the dual analysis) from each volunteer. The breath samples were characterization of the breath samples with an array of sensors, combined with a statistical pattern recognition algorithm, with the aim of identifying specific patterns (the so-called breath prints) for GC and nonmalignant gastric conditions.

Samples are collected through stainless steel chamber of 140 ml volume which is standardized. Samples could be collected even from the elderly or bedridden patients without causing discomfort. The collection period was 1.5 min at 0.5 per 1 min, and the dead space of samples is removed by setting the system timer. Volunteers are from various hospitals and dispensary in and around the Tirunelveli district, Tamil Nadu, India. All subject samples were collected in random order, sample collection from subject volunteer is shown in Figure 2.
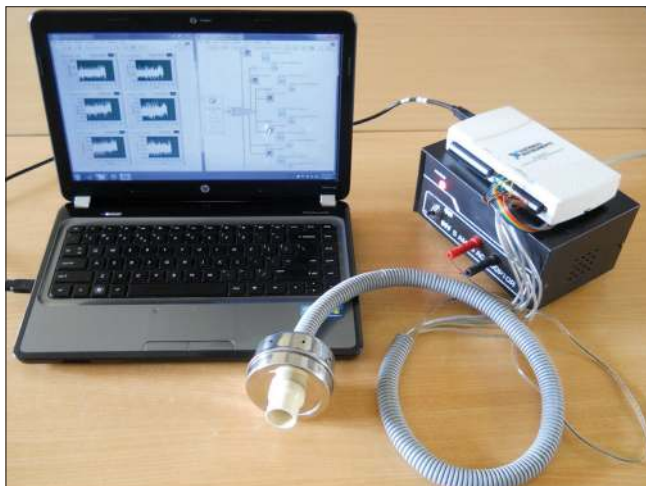
Breath samples were collected after written informed consent from 270 volunteers, aged 21–73 years, at the CSI Jayaraj Annapackiam Mission Hospital and CSI Bell Pins Inndrani Chelladurai Mission Hospital, Palayamkottai, Tirunelveli. All volunteers underwent upper digestive endoscopy after recruitment according to the hospital's routine clinical protocol. Biopsy samples were taken for histopathology if lesions (including ulceration of the stomach lining) were visually observed.

The following exclusion criteria were applied before sample collection: patients who have undergone gastric resection in the past; patients who were found to suffer from endoscopically detectable precancerous conditions (e.g., mucosal atrophy); and patients who took medication affecting gastric acid secretion (e.g., proton pump inhibitors) and/or antibiotics during an interval of 1 month before the breath test. The reason for the latter exclusion criterion for this study was that previous medication could strongly affect the composition of the exhaled breath.

After excluding, we employed the breath samples of 161 patients were analyzed for this study: 49 GC patients, 19 patients with benign gastric ulcers, and 11 patients with less severe gastric conditions are shown in Table 1. The less severe stomach conditions cases included with no endoscopic abnormalities (82) and with endoscopic abnormalities without ulceration (11).

Ethical approval was obtained from the Ethics Committee of Periyar University, Salem, Tamil Nadu, India, and the clinical trial was registered. The treatment decisions were based solely on the conventional diagnosis described above.

Table 1: Composition of the subject database

Type of subject	Number	Gender (male/female)	Age	Diagnosis
Healthy	82	34/48	20-32	
Abnormalities without ulceration	11	9/2	24-35	Endoscopy only
Gastric ulcer	19	11/8	45-58	Endoscopy
Gastric cancer	49	25/24	38-55	with Biopsy

Figure 1: Electronic-nose system using data acquisition card

Figure 2: Collection of breath samples from cancer volunteers

Neither the patients nor their treating physicians were informed of the results of the breath tests.

EXPERIMENTAL RESULTS

Data Sampling

The dataset consists of experimentally obtained 161 observations. The dataset was divided into ten disjoint subsets, namely, training set containing 145 observations (90% of total observations) and test set comprising 16 observations (10% of total observations). The overall acceptability was used as output parameter for developing the artificial intelligence tool.

Artificial Neural Networks

Cascade-forward back-propagation (CFBP) and feed-forward back-propagation (FFBP) artificial intelligence models [Figures 3-6] were trained with the breath sampled dataset. Different combinations of several internal parameters, i.e., data preprocessing, data partitioning approach, number of hidden layers, number of neurons in each hidden layer, transfer function, error goal, etc., were attempted. Different variants of the back-propagation algorithm were tried here: Levenberg-Marquardt (LM), Bayesian regularization, BFGS Quasi-Newton, Resilient (RP), scaled conjugate gradient, conjugate gradient with Powell/Beale restarts, conjugate gradient with Fletcher-Powell, conjugate gradient with Polak-Ribiére, one step secant, variable learning rate gradient descent, gradient descent with momentum, gradient descent shown in Tables 2-5 with different activation functions such as radial basis, normalized radial basis, triangular basis, hyperbolic tangent sigmoid, Elliot symmetric sigmoid, Elliot 2 symmetric sigmoid, hard-limit, symmetric hard-limit, competitive, soft max shown in Tables 6-9. RP algorithm produced better results, during training of single- and multi-layer FFBP neural network as shown in Figures 7-12; however, during training of multilayer
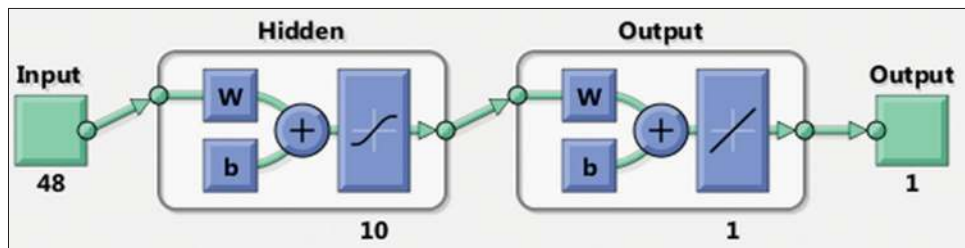

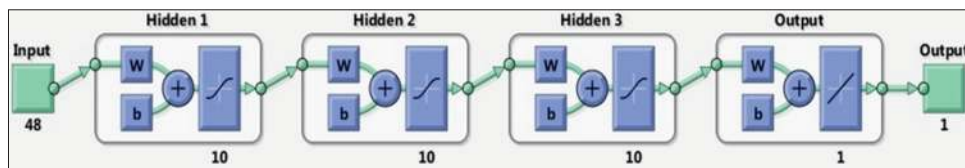**Figure 3:** Network diagram of single-layer feed-forward neural network classifiers

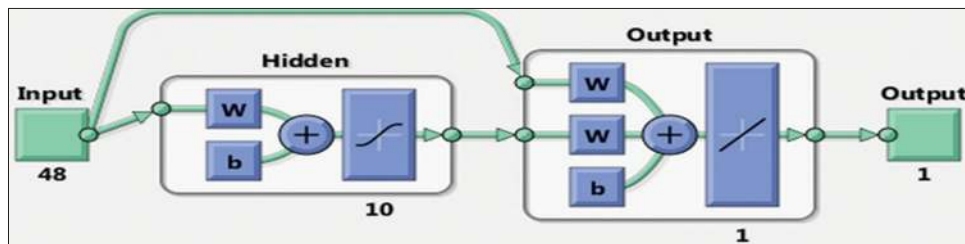**Figure 4:** Network diagram of multilayer feed-forward neural network classifiers

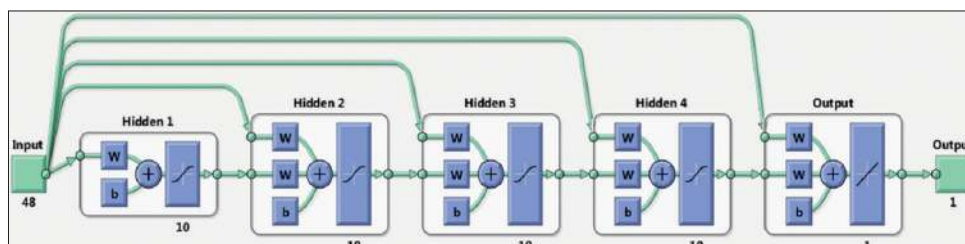**Figure 5:** Network diagram of single-layer cascade-forward neural network classifiers

Figure 6: Network diagram of multilayer cascade-forward neural network classifiers

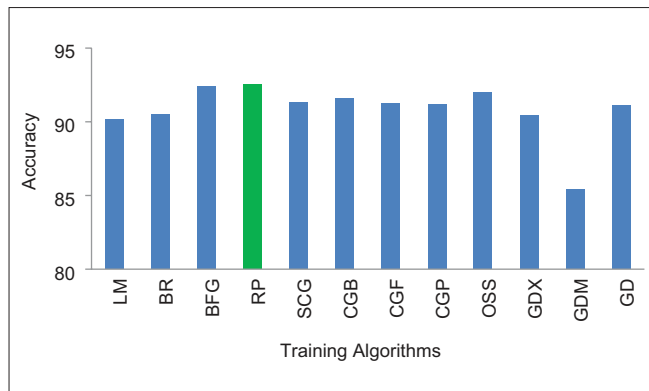## Table 2: Classification accuracy of single-layer feed-forward neural network

Training algorithm	Percentage	MSE	MAE	SSE	SAE
LM	90.14	0.124	0.235	1.867	3.758
BR	90.47	0.096	0.192	1.543	3.062
BFG	92.39	0.09	0.185	1.253	2.975
RP	92.54	0.086	0.203	1.378	3.252
SCG	91.3	0.089	0.208	1.422	3.326
CGB	91.58	0.087	0.209	1.387	3.341
CGF	91.25	0.09	0.211	1.442	3.379
CGP	91.17	0.09	0.209	1.43	3.356
OSS	92.01	0.098	0.198	1.325	3.188
GDX	90.43	0.108	0.226	1.628	3.616
GDM	85.39	0.226	0.332	3.623	5.317
GD	91.09	0.09	0.214	1.443	3.418
Average	90.81	0.095	0.198	1.489	3.185

MSE – Mean squared normalized error; MAE – Mean absolute error; SSE – Sum squared error; SAE – Sum absolute error; LM – Levenberg-Marquardt algorithm; BR – Bayesian regularization; SCG – Scaled conjugate; CGB – Conjugate gradient with Powell/Beale restarts; CGF – Conjugate gradient with Fletcher-Powell; CGP – Conjugate gradient with Polak-Ribiére; OSS – One step secant; GDM – Gradient descent with momentum; GD – Gradient descent; GDX – Variable learning rate gradient descent; RP – Resilient; BFG – BFGS Quasi-Newton

Table 4: Classification accuracy of single-layer cascade-forward neural network

Training algorithm	Percentage	MSE	MAE	SSE	SAE
LM	92.25	0.1	0.228	1.604	3.655
BR	92.39	0.087	0.186	1.185	2.969
BFG	92.26	0.092	0.222	1.465	3.569
RP	92.49	0.104	0.244	1.67	3.897
SCG	92.06	0.093	0.23	1.486	3.687
CGB	92.3	0.089	0.222	1.436	3.547
CGF	92.16	0.098	0.224	1.459	3.59
CGP	92.1	0.089	0.222	1.418	3.551
OSS	91.78	0.099	0.236	1.586	3.773
GDX	92.14	0.092	0.224	1.452	3.584
GDM	82.88	0.919	0.638	14.71	10.21
GD	92.03	0.096	0.231	1.539	3.693
Average	91.4	0.168	0.261	2.673	4.188

MSE – Mean squared normalized error; MAE – Mean absolute error; SSE – Sum squared error; SAE – Sum absolute error; LM – Levenberg-Marquardt algorithm; BR – Bayesian regularization; SCG – Scaled conjugate; CGB – Conjugate gradient with Powell/Beale restarts; CGF – Conjugate gradient with Fletcher-Powell; CGP – Conjugate gradient with Polak-Ribiére; OSS – One step secant; GDM – Gradient descent with momentum; GD – Gradient descent; GDX – Variable learning rate gradient descent; RP – Resilient; BFG – BFGS Quasi-Newton

Table 3: Classification accuracy of multilayer feed-forward neural network

Training algorithm	Percentage	MSE	MAE	SSE	SAE
LM	91.42	0.094	0.186	1.503	2.974
BFG	93.04	0.071	0.161	1.136	2.581
RP	93.17	0.077	0.153	1.111	2.444
SCG	92.09	0.083	0.197	1.325	3.104
CGB	92.94	0.072	0.173	1.158	2.763
CGF	92.03	0.083	0.19	1.33	3.038
CGP	92.41	0.076	0.176	1.224	2.822
OSS	92.36	0.079	0.181	1.27	2.892
GDX	91.24	0.093	0.21	1.484	3.357
GDM	87.83	0.161	0.289	2.583	4.621
GD	91.58	0.085	17.21	1.372	3.185
Average	91.83	0.088	0.33	1.408	3.071

MSE – Mean squared normalized error; MAE – Mean absolute error; SSE – Sum squared error; SAE – Sum absolute error; LM – Levenberg-Marquardt algorithm; BR – Bayesian regularization; SCG – Scaled conjugate; CGB – Conjugate gradient with Powell/Beale restarts; CGF – Conjugate gradient with Fletcher-Powell; CGP – Conjugate gradient with Polak-Ribiére; OSS – One step secant; GDM – Gradient descent with momentum; GD – Gradient descent; GDX – Variable learning rate gradient descent; RP – Resilient; BFG – BFGS Quasi-Newton

Table 5: Classification accuracy of multilayer cascade-forward neural network

Training algorithm	Percentage	MSE	MAE	SSE	SAE
LM	92.62	0.089	0.213	1.432	3.433
BFG	92.35	0.086	0.213	1.381	3.405
RP	92.38	0.114	0.25	1.816	4.009
SCG	92.21	0.097	0.223	1.419	3.571
CGB	92.34	0.091	0.22	1.454	3.518
CGF	92.29	0.09	0.224	1.412	3.567
CGP	92.18	0.09	0.225	1.446	3.599
OSS	91.99	0.095	0.23	1.512	3.68
GDX	92.13	0.098	0.238	1.562	3.805
GDM	91.91	0.281	0.309	4.823	5.099
GD	91.91	0.093	0.227	1.491	3.629
Average	92.21	0.111	0.233	1.795	3.755

MSE – Mean squared normalized error; MAE – Mean absolute error; SSE – Sum squared error; SAE – Sum absolute error; LM – Levenberg-Marquardt algorithm; BR – Bayesian regularization; SCG – Scaled conjugate; CGB – Conjugate gradient with Powell/Beale restarts; CGF – Conjugate gradient with Fletcher-Powell; CGP – Conjugate gradient with Polak-Ribiére; OSS – One step secant; GDM – Gradient descent with momentum; GD – Gradient descent; GDX – Variable learning rate gradient descent; RP – Resilient; BFG – BFGS Quasi-Newton

Figure 7: Resilient back-propagation training algorithm outperformance: good in classification accuracy of single-layer feed-forward neural network

Figure 8: Resilient back-propagation training algorithm outperformance: good in error rate of single-layer feed-forward neural network

Table 6: Classification accuracy of single-layer feed-forward neural network based on activation functions
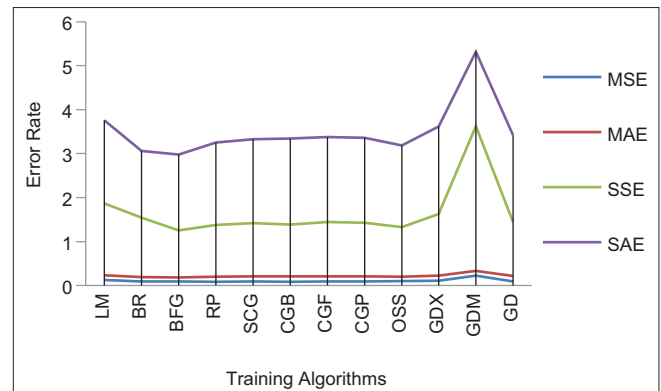
Activation function	Percentage	MSE	MAE	SSE	SAE
RBN	90.8	0.107	0.22	1.537	3.514
NRBN	90	0.106	0.226	1.688	3.599
TBN	90.97	0.1	0.218	1.605	3.52
HTSN	90.88	0.1	0.217	1.587	3.465
ESSN	90.76	0.1	0.217	1.594	3.477
E2SSN	91.03	0.124	0.224	1.903	3.587
HLN	90.85	0.108	0.219	1.633	3.504
SHLN	91.71	0.093	0.203	1.433	3.246
CN	90.45	0.113	0.219	1.661	3.508
SMN	90.68	0.113	0.223	1.808	3.571
Average	90.81	0.106	0.219	1.645	3.499

MSE – Mean squared normalized error; MAE – Mean absolute error; SSE – Sum squared error; SAE – Sum absolute error; RBN – Radial basis network; NRBN – Normalized radial basis network; TBN – Triangular basis network; HTSN – Hyperbolic tangent sigmoid network; ESSN – Elliot symmetric sigmoid network; E2SSN – Elliot 2 symmetric sigmoid network; HLN – Hard-limit network; SHLN – Symmetric hard-limit network; CN – Competitive network; SMN – Soft max network

## Table 8: Classification accuracy of single-layer cascade-forward neural network based on activation functions

Activation function	Percentage	MSE	MAE	SSE	SAE
RBN	91.33	0.19	0.271	3.023	4.339
NRBN	91.49	0.154	0.254	2.38	4.059
TBN	91.35	0.151	0.256	2.414	4.098
HTSN	91.36	0.123	0.249	1.963	3.98
ESSN	91.56	0.211	0.269	3.366	4.297
E2SSN	91.24	0.19	0.266	2.926	4.255
HLN	91.4	0.146	0.254	2.329	4.061
SHLN	91.5	0.137	0.248	2.181	3.963
CN	91.26	0.162	22.08	2.585	4.188
SMN	91.56	0.168	0.262	2.672	4.198
Average	91.40	0.163	2.441	2.584	4.144

MSE – Mean squared normalized error; MAE – Mean absolute error; SSE – Sum squared error; SAE – Sum absolute error; RBN – Radial basis network; NRBN – Normalized radial basis network; TBN – Triangular basis network; HTSN – Hyperbolic tangent sigmoid network; ESSN – Elliot symmetric sigmoid network; E2SSN – Elliot 2 symmetric sigmoid network; HLN – Hard-limit network; SHLN – Symmetric hard-limit network; CN – Competitive network; SMN – Soft max network

Table 7: Classification accuracy of multilayer feed-forward neural network based on activation functions

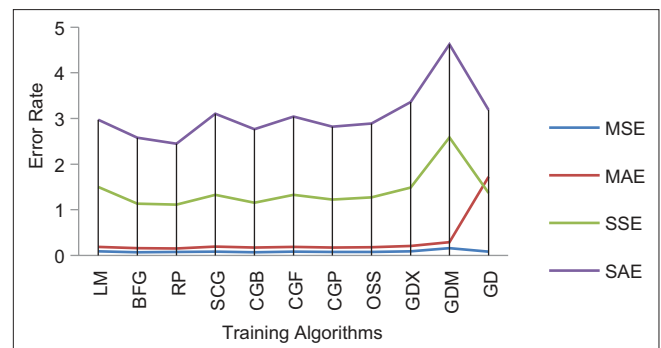| Activation function	Percentage	MSE	MAE	SSE	SAE
RBN	92.09	0.084	0.173	1.255	2.77
NRBN	92.15	0.077	0.172	1.23	2.754
TBN	91.84	0.082	0.179	1.304	2.864
sHTSN	91.62	0.081	0.176	1.293	2.823
ESSN	91.58	0.081	0.179	1.299	2.82
E2SSN	91.83	0.079	0.177	1.257	2.832
HLN	91.75	0.081	0.175	1.292	2.8
SHLN	91.77	0.08	0.175	1.277	2.798
CN	91.79	0.094	0.187	1.498	2.996
SMN	90.68	0.113	0.223	1.808	3.571
Average	91.82	0.082	0.182	1.291	2.815

MSE – Mean squared normalized error; MAE – Mean absolute error; SSE – Sum squared error; SAE – Sum absolute error; RBN – Radial basis network; NRBN – Normalized radial basis network; TBN – Triangular basis network; HTSN – Hyperbolic tangent sigmoid network; ESSN – Elliot symmetric sigmoid network; E2SSN – Elliot 2 symmetric sigmoid network; HLN – Hard-limit network; SHLN – Symmetric hard-limit network; CN – Competitive network; SMN – Soft max network
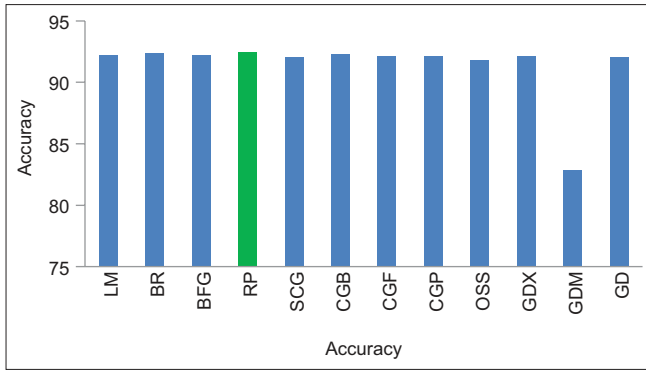
Table 9: Classification accuracy of multilayer cascade-forward neural network based on activation functions

Activation function	Percentage	MSE	MAE	SSE	SAE
RBN	92.24	0.087	0.204	1.658	3.378
NRBN	92.05	0.094	0.214	1.501	3.417
TBN	92.27	0.153	0.234	2.446	3.749
HTSN	92.2	0.088	0.211	1.408	3.373
ESSN	92.23	0.098	0.216	1.46	3.454
E2SSN	92.32	0.119	0.217	1.895	3.478
HLN	92.15	0.088	0.208	1.395	3.322
SHLN	92.18	0.119	0.218	1.896	3.506
CN	92.36	0.085	0.207	1.33	3.306
SMN	92.11	0.092	0.215	1.467	3.447
Average	92.21	0.102	0.214	1.646	3.443

MSE – Mean squared normalized error; MAE – Mean absolute error; SSE – Sum squared error; SAE – Sum absolute error; RBN – Radial basis network; NRBN – Normalized radial basis network; TBN – Triangular basis network; HTSN – Hyperbolic tangent sigmoid network; ESSN – Elliot symmetric sigmoid network; E2SSN – Elliot 2 symmetric sigmoid network; HLN – Hard-limit network; SHLN – Symmetric hard-limit network; CN – Competitive network; SMN – Soft max network
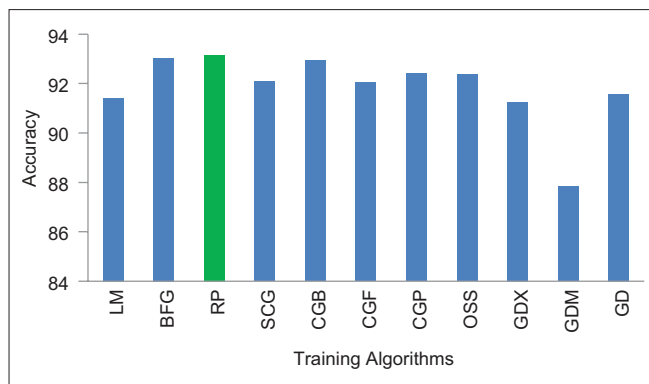
Figure 9: Resilient back-propagation training algorithm outperformance: good in classification accuracy of multilayer feed-forward neural network

Figure 10: Resilient back-propagation training algorithm outperformance: good in error rate of multilayer feed-forward neural network

CFBP neural network, LM produces better results as shown in Figures 13 and 14. Normalized radial basis, symmetric hard-limit, competitive, soft max activation function supports to

the algorithms outperformance are shown in Figures 15-22. There is no generalized method to determine the optimum values for number of hidden layers, neurons in each hidden layer, etc., as they are working of expected intelligence.
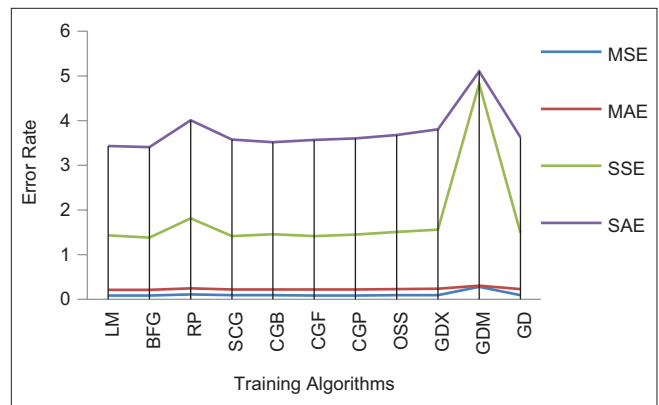
Figure 11: Resilient back-propagation training algorithm outperformance: good in classification accuracy of single-layer cascade-forward neural network
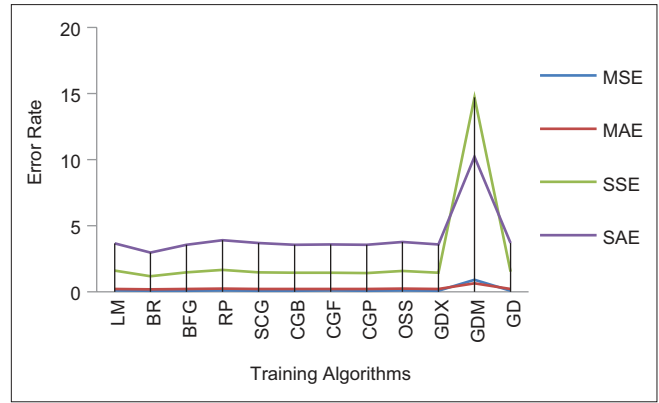

Figure 12: Resilient back-propagation training algorithm outperformance: good in error rate of single-layer cascade-forward neural network
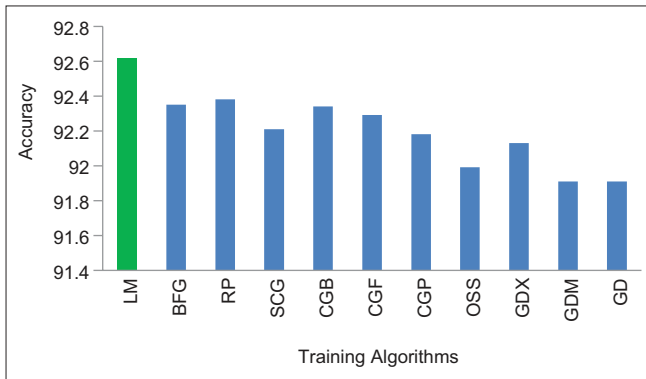

Figure 13: Levenberg-Marquardt back-propagation training algorithm outperformance: good in classification accuracy of multilayer cascade-forward neural network

Figure 14: Levenberg-Marquardt back-propagation training algorithm outperformance: good in error rate of multilayer cascade-forward neural network
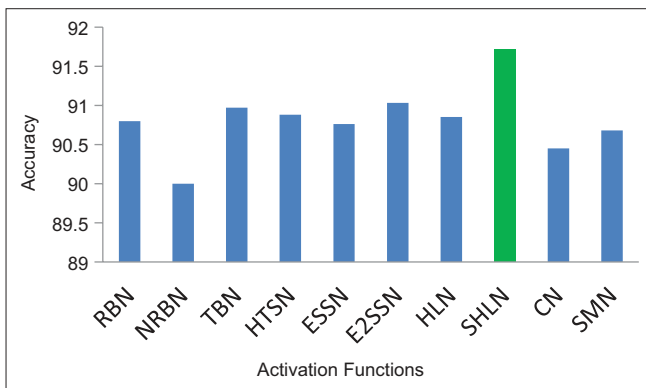
Figure 15: Symmetric hard-limit activation function outperformance: good in classification accuracy of single-layer feed-forward neural network
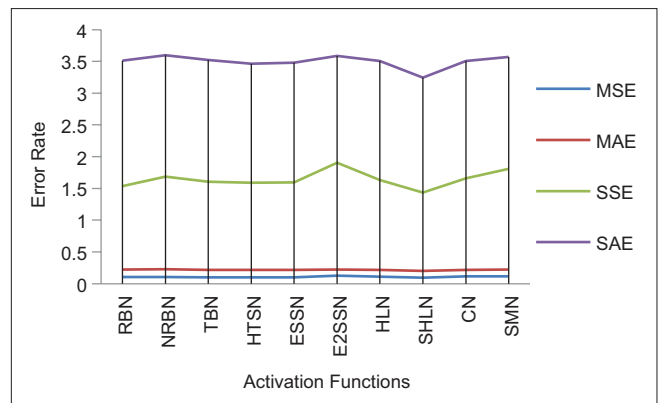
Figure 16: Symmetric hard-limit activation function outperformance: good in error rate of single-layer feed-forward neural network

Initially, the reduced feature set selected from the feature selection methods is normalized between zero and one. That is each value in the feature set is divided by the maximum value from the set. These normalized values are assigned to the input neurons.

The number of hidden neurons is greater than or equal to the number of input neurons. Moreover, there is only one output neuron. Initial weights are assigned randomly. The
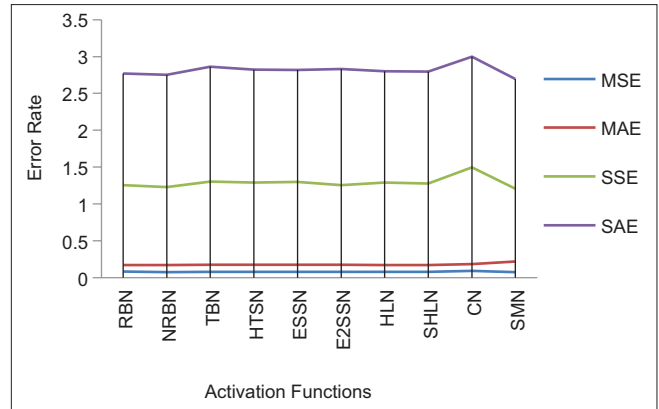
output from each hidden neuron is calculated using the sigmoid function:

$$S_1 = \frac{1}{1+e^{-\lambda x}} \text{ , where } \lambda = 1 \text{ and } x = \sum_i w_{ih} k_i \qquad (1)$$

where w_{ih} is the weight assigned between input and hidden layer and k is the input value. The output
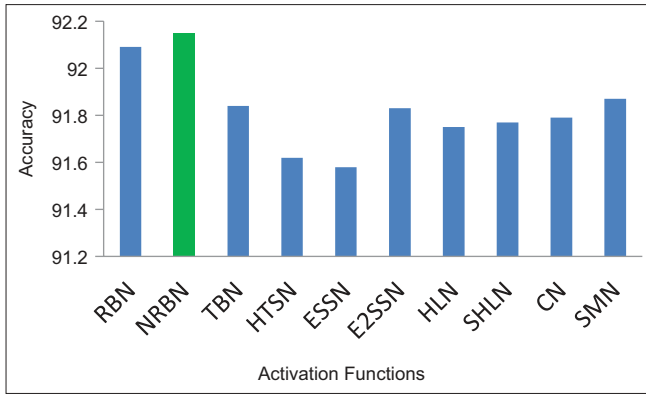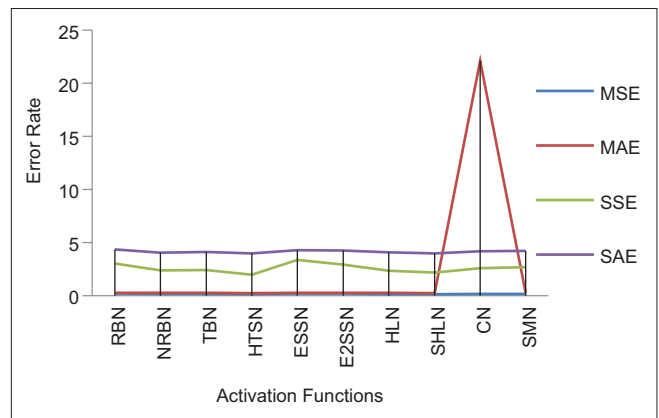
Figure 17: Normalized radial basis activation function outperformance: good in classification accuracy of multilayer feed-forward neural network


Figure 18: Normalized radial basis activation function outperformance: good in error rate of multilayer feed-forward neural network
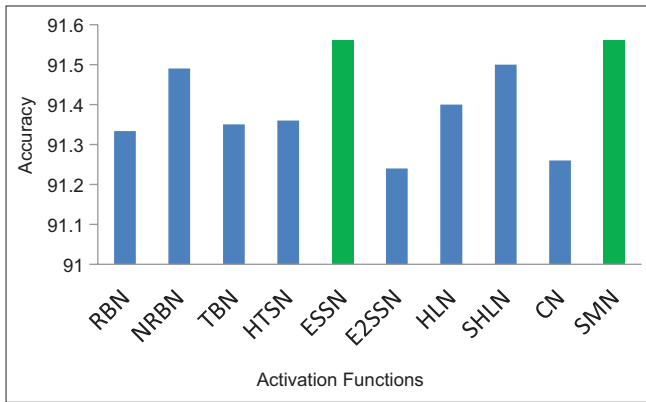

Figure 19: Elliot symmetric sigmoid and soft max activation function outperformance: good in classification accuracy of single-layer cascade-forward neural network

Figure 20: Elliot symmetric sigmoid and soft max activation function outperformance: good in error rate of single layer cascade-forward neural network
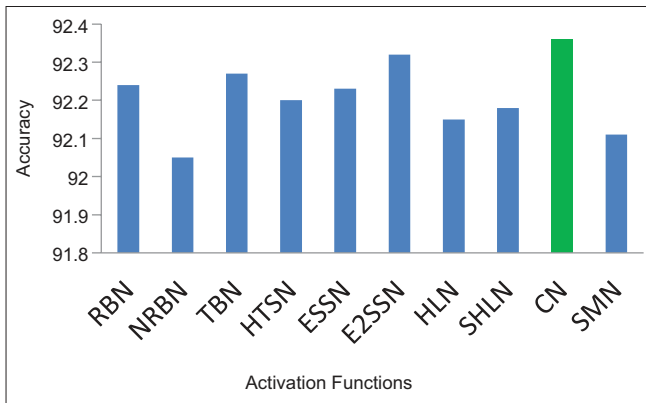
Figure 21: Competitive activation function outperformance: good in classification accuracy of multilayer cascade-forward neural network
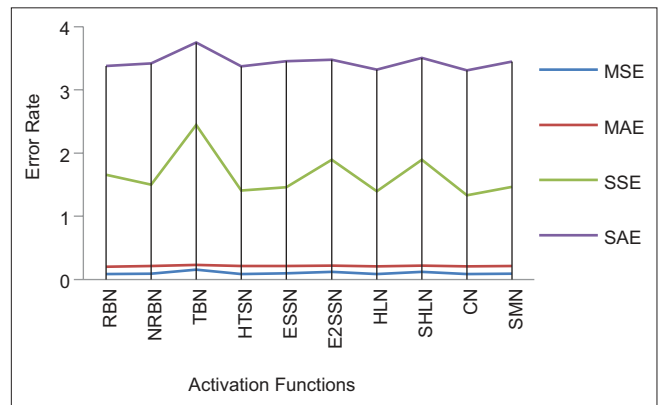
Figure 22: Competitive activation function outperformance: good in error rate of multilayer cascade-forward neural network

from the output layer is calculated using the sigmoid function.

$$S_2 = \frac{1}{1+e^{-\lambda x}} \text{ , where } \lambda = 1 \text{ and } x = \sum_i w_{ho} S_i \qquad (2)$$

where w_{ho} is the weight assigned between hidden and output layer and S_i is the output value from hidden neurons. S_2 is subtracted from the desired output.

Using this error (e) value, the updating of weight is performed as:

$$\delta = eS_2 (1 - S_2) \qquad (3)$$

The weights assigned between the input and the hidden layer and the hidden and output layer are updated as:

$$w_{ho} = w_{ho} + (n\delta\, S_1) \qquad (4)$$

$$w_{ih} = w_{ih} + (n\delta\, k_i) \qquad (5)$$

where n is the learning rate and k is the input value. Again the output is calculated from the hidden and output neurons. Then, the error (e) value is checked and the weights get updated.[2] This procedure is repeated till the target output is equal to the desired output. The algorithm of back-propagation classifier for classification is shown below.[10]

Back-propagation neural network (C, D)
   C, Conditional features; D, Decision feature
1. Normalize the feature between 0 and 1 and assigned to input neurons
2. Initialize random weight
3. Compute each hidden neuron, $S_1 = \dfrac{1}{1+e^{-\lambda x}}$, $x = \sum_i w_{ih} k_i$
4. Compute output layer, $S_2 = \dfrac{1}{1+e^{-\lambda x}}$, $x = \sum_i w_{ho} S_i$
5. Subtract S_2 from the desired output. Using error (e) value, compute weight as: $\delta = eS_2(1-S_2)$
6. Update the weights using this δ value
7. $w_{ho} = w_{ho} + (n\delta\, S_1)$
8. $w_{ih} = w_{ih} + (n\delta\, k_i)$
Perform steps 4 to 8 with the updated weights, till the target output is equal to the desired output

---

Feed-forward back-propagation model

FFBP artificial intelligence model consists of input, hidden, and output layers. Back-propagation learning algorithm was used for learning these networks. During training this network, calculations were carried out from input layer of network toward output layer, and error values were then propagated to prior layers. Feed-forward networks often have one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons. Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors. The linear output layer lets the network produce values outside the range −1 to +1. On the other hand, outputs of a network such as between 0 and 1 are produced, then the output layer should use a sigmoid transfer function.[11]

Cascade-forward back-propagation model

CFBP models are similar to feed-forward networks but include a weight connection from the input to each layer and from each layer to the successive layers. While two-layer feed-forward networks can potentially learn virtually any input-output relationship, feed-forward networks with more layers might learn complex relationships more quickly. For example, a three-layer network has connections from layer 1 to layer 2, layer 2 to layer 3, and layer 1 to layer 3. The three-layer network also has connections from the input to all three layers. The additional connections might improve the speed at which the network learns the desired relationship.[12] CFBP artificial intelligence model is similar to FFBP neural network in using the back-propagation algorithm for weights updating, but the main symptom of this network is that each layer of neurons related to all previous layer of neurons.[11]

The performance of CFBP and FFBP were evaluated using mean squared normalized error, mean absolute error, sum squared error, and sum absolute error technique.

The functionality of 12 different training algorithms, which are used in this work, is synopsized in Table 10. A short description of all training algorithms is presented in Table

Table 10: Description of artificial neural networks training algorithms

Description
Levenberg-Marquardt algorithm[13,14]
Bayesian regularization[15]
BFGS Quasi-Newton[16]
Resilient[17]
Scaled conjugate[18]
Conjugate gradient with Powell/Beale restarts[19-21]
Conjugate gradient with Fletcher-Powell[19,22]
Conjugate gradient with Polak-Ribiére[19,21]
One step secant[23]
Variable learning rate gradient descent[24]
Gradient descent with momentum[24]
Gradient descent[24]

	SLFF	MLFF	SLCF	MLCF
SAE	3.185	3.071	4.188	3.755
SSE	1.489	1.408	2.673	1.795
MAE	0.198	0.33	0.261	0.233
MSE	0.095	0.088	0.168	0.111
Accuracy	90.81	91.82	91.4	92.21

Figure 23: Classification accuracy outperformance in different neural network based on algorithms

	SLFF	MLFF	SLCF	MLCF
SAE	3.499	2.815	4.144	3.443
SSE	1.645	1.291	2.584	1.646
MAE	0.219	0.182	2.441	0.214
MSE	0.106	0.082	0.163	0.102
Accuracy	90.81	91.82	91.4	92.21

Figure 24: Classification accuracy outperformance in different neural network based on activation function

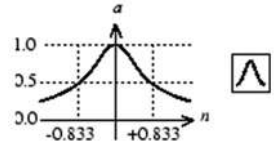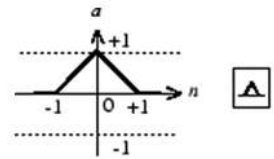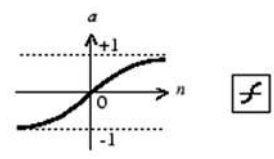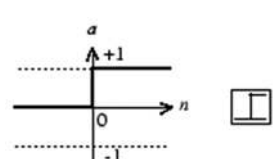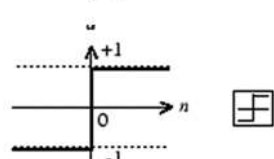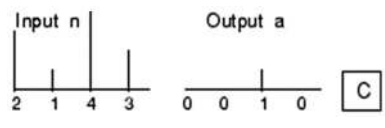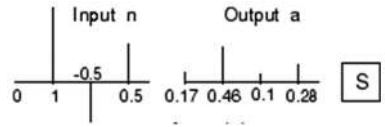10[25] while more analytical representations are shown in Table 10.[13-24] The basic steps of the back-propagation algorithm have been described in several textbooks.[26,27] The functionality of ten different activation functions, which are used in this work, is synopsized in Table 11.[28,29] The overall performance of the neural network based on the algorithm and activation functions are shown in Figures 23 and 24.

CONCLUSION

In this study, neural network has been used to classify the GC as malignant or benign or normal. Based on the obtained results, the RP algorithm produced, up to the mark of classification accuracy (92.54%, 93.17%, and 92.49%), during training of single- and multi-layer FFBP neural network; however, by multilayer CFBP neural network, LM (92.62%) produces better classification accuracy. Normalized radial basis (92.15%), symmetric hard-limit (91.71%), Elliot symmetric sigmoid (91.56%), competitive (92.36%), soft max (91.56%) activation function supports the algorithms performance for the better classification. It was also observed that in general, multi hidden layer network provided the better classification accuracy compared to the single hidden layer network to classify the breath samples of GC. In the near future, we need to standardize the procedures and develop a learning system widely acceptable by breath analysts worldwide. In this way, we will be able to reduce deaths due to GC, the second leading cause of cancer deaths worldwide.

Table 11: Description of artificial neural networks activation functions

Activation function	Input/output relation	Icon
Radial basis	$\exp(-n^2)$	
Normalized radial basis	$\exp(-n^2)/\text{sum}(\exp[-n^2])$	
Triangular basis	1 - abs (n), if $-1 \le n \le 1$ 0, otherwise	
Hyperbolic tangent sigmoid	$2/(1+\exp[-2 \times n]) - 1$	
Elliot symmetric sigmoid	$(x \times n)/(1 + |x \times n|)$	
Elliot 2 symmetric sigmoid	$n \times 1/([1 + |x \times n|] \times [1 + |x \times n|])$	
Hard-limit	1, if $n \ge$ 0, otherwise	
Symmetric hard-limit	1 if $n \ge$ 0, -1 otherwise	
Competitive	1, neuron with max n 0, all other neurons	
Soft max	$\exp(n)/\text{sum}(\exp[n])$	

Conflicts of Interest

There are no conflicts of interest.

REFERENCES

1. Altay Güvenir H, Emeksiz N, Ikizler N, Ormeci N. Diagnosis of gastric carcinoma by classification on feature projections. Artif Intell Med 2004;31:231-40.

2. Hryniuk A, Ross BM. A preliminary investigation of exhaled breath from patients with celiac disease using selected ion flow tube mass spectrometry. J Gastrointestin Liver Dis 2010;19:15-20.

3. Smith D, Spanel P. The challenge of breath analysis for clinical diagnosis and therapeutic monitoring. Analyst 2007;132:390-6.

4. Dikshit R, Gupta PC, Ramasundarahettige C, Gajalakshmi V, Aleksandrowicz L, Badwe R, et al. Cancer mortality in India: A nationally representative survey. Lancet 2012;379:1807-16.

5. Kalyani R, Das S, Kumar ML. Pattern of cancer in adolescent and young adults – A ten year study in India. Asian Pac J Cancer Prev 2010;11:655-9.

6. Sarker SK, Sinha VK, Chaudhry R, Maudar KK. Gastric cancer: A critical analysis of surgical treatment and long term survival. J Indian Med Assoc 1992;90:61-4.

7. Kajitani T. The general rules for the gastric cancer study in surgery and pathology. Part I. Clinical classification. Jpn J Surg 1981;11:127-39.

8. Kurihara H. Detection of early gastric cancer outside the mass screening program. Jpn J Clin Oncol 1998;28:233.

9. Yokota T. Bormann's type IV gastric cancer – Clinicopathologic analysis. Can J Surg 1999;42:371-6.

10. Freeman JA, Skapura DM. Neural Networks Algorithms, Applications and Programming Techniques. USA: Addison-Wesley Pub (Sd) ; 1991. p. 115-6.

11. Chayjan RA. Modeling of sesame seed dehydration energy requirements by a soft-computing. Aust J Crop Sci 2010;4:180-4.

12. Demuth H, Beale M, Hagan M. Neural Network Toolbox User's Guide. Natick, USA: The MathWorks Inc.; 2009.

13. Levenberg K. A method for the solution of certain problems in least squares. Q Appl Math 1944;2:164-8.

14. Marquardt D. An algorithm for least squares estimation of nonlinear parameters. SIAM J Appl Math 1963;11:431-41.

15. Aggarwal KK, Singh Y, Chandra P, Puri M. Bayesian regularization in a neural network model to estimate lines of code using function points. J Comput Sci 2005;2005:505-9.

16. Battiti R. First and second order methods for learning – Between steepest descent and Newton's method. Neural Comput 1992;4:141-66.

17. Riedmiller M, Braun H. A Direct Adaptive Method for Faster Backpropagation Learning. The RPROP Algorithm. Proceedings of the IEEE International Conference on Neural Networks. Vol. 1. Proceedings of the IEEE International Conference on Neural Networks, San Francisco; March, 1993. p. 586-91.

18. Moller MF. A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw 1993;6:525-33.

19. Saini LM, Soni MK. Artificial neural network-based peak load forecasting using conjugate gradient methods. IEEE Trans Power Syst 2002;17:907-12.

20. Powell MJ. Restart procedures for the conjugate gradient method. Math Program 1977;12:241-54.

21. Polak E. Computational Methods in Optimization – A Unified Approach. New York: Academic Press; 1971.

22. Fletcher R, Reeves CM. Function minimization by conjugate gradients. Comput J 1964;7:149-54.

23. Shaharin R, Prodhan UK, Rahman M. Performance study of TDNN training algorithm for speech recognition. Int J Adv Res Comput Sci Technol 2014;2:90-5.

24. Ghosh PS, Chakravorti S, Chatterjee N. Estimation of time-to-flashover characteristics of contaminated electrolytic surfaces using a neural network. IEEE Trans Dielectr Electr Insul 1995;2:1064-74.

25. Asimakopoulou GH, Kontargyri VT, Tsekouras GJ, Asimakopoulou FH, Gonos IF, Stathopulos IA. Artificial neural network optimization methodology for the estimation of the critical flashover voltage on insulators. IET Sci Meas Technol 2008;3:90-104.

26. Haykin S. Neural Networks – A Comprehensive Foundation. New Jersey: Englewood Cliffs, NJ, Prentice Hall; 1998.

27. Tsekouras GJ, Koukoulis J, Nikolinakou MA, Mastorakis NE. Prediction of Face Settlement During Tunneling Excavation using Artificial Neural Network. WSEAS International Conference on Engineering Mechanics, Structures, Engineering Geology (EMESEG '08). Heraklion, Crete Island, Greece; July, 2008. p. 22-5.

28. Karlik B, Olgac V. Performance analysis of various activation functions in generalized MLP architectures of neural networks. Int J Artif Intell Expert Syst 2010;1:111-22.

29. Sibi P, Jones SA, Siddarth P. Analysis of different activation functions using back propagation neural networks. J Theor Appl Inf Technol 2013;47:1264-8.