Ensembling of EGFR Mutations' based Artificial Neural Networks for Improved Diagnosis of Non-Small Cell Lung Cancer

Emmanuel ADETIBA
Dept. of Electrical & Information Eng., CST,
Covenant University, Ota, Nigeria.

Frank A. IBIKUNLE
Dept. Electrical and Information Eng., CST,
Covenant University, Ota, Nigeria.

ABSTRACT

In this research work, we built and ensembled different EGFR microdeletion mutations' based Artificial Neural Networks(ANNs) for improved diagnosis of Non-Small Cell Lung Cancer(NSCLC). We developed two novel algorithms, namely; Genomic Nucleotide Encoding & Normalization (GNEN) algorithm to encode and normalize the EGFR nucleotides and SimMicrodel algorithm to programmatically simulate microdeletion mutations. Sample patients' data with microdeletion mutations were extracted from online EGFR mutation databases and the two novel algorithms (implemented in MATLAB) were applied to these data to generate appropriate data sets for training and testing of the networks.

The networks after proper training, were combined using minimum error voting ensembling to predict the number of nucleotide deletions in NSCLC patients. Using this ensembling approach, our simulations achieved predictions with minimal error and provides a basis for diagnosing NSCLC patients using genomics based ANN.

Key Words: ANN, EGFR, GNEN, NSCLC, LM, SimMicrodel,

1 INTRODUCTION

Environmental and genetic factors play vital roles in the deveopment of any disease. However, certain human diseases are categorised as genetic disease or disorder because they are caused by abnormalities in an individual's genetic materials called genome. This class of disease are of four different types which are; single-gene, multifactorial, chromosomal and mitochondrial. Meanwhile, the normal function of a gene is to encode a protein not to cause illness but genetic diseases occur when genes are unable to work properly. Cancer in humans generally arise from alterations in oncogenes, tumor suppressor gene or genes whose products participate in genome surveillance[1]. It can be considered a multifactorial disease because it results from the combined influence of many genetic factors acting in concert with environmental insults such as ultraviolet radiation, cigarette smoke and viruses.

Some chemicals that are used or released via industrial activities such as production of plastics, asbestos, pharmacautical products and food supplements are not only mutagens but carcinogenic(cancer producing). For instance, chemicals that are released in smokes from cigarette imparts a large number of particles on the airways and alveoli of the human lung which slowly oxidize amd produce genotoxic radicals. A large percentage of lung cancers which is characterised by an uncontrollable growth of cells in the lungs are attributable to cigaratte smoking. About 85% of people who develop lung cancer are either smokers or have been smoker. Worlwid in the 21st century, lung cancers emerged as the leading cause of cancer deaths because it results in an estimated 1.3 million deaths each year[1,2,3].

Pathologists determine the type of lung cancer by looking at a biopsy of tumor cells under the microscope. There are two major types of lung cancer which are; non-small cell lung cancer(NSCLC) and small cell lung cancer(SCLC). NSCLC accounts for about 85% of lung cancers while the remaining 15% are SCLC[3,4]. A spectrum of mutations exists within the EGFR kinase domain in tumours of patients with NSCLC. The most frequently observed mutations are the

Exon 19 deletions and the Exon 21 L858R mutation, which taken together account for approximately 85% - 90% of all EGFR mutations[5,6,7]. Also, experimental studies reported in [8] have shown both qualitative and quantitative alterations in downstream signalling by mutant EGFR and suggested that NSCLC cells with these mutations may be dependent on the altered signals for survival.

The various mutational patterns in EGFR is a strong basis for an electronic based diagnosis of NSCLC using artificial neural network (ANN). The use of ANN in biological and medical researches has proliferated greatly during the last few years. ANN attempts to emulate function of the human brain and has played a great role in the fields of cancer research for diagnosis, prognosis and management of stages[12].

Hansen et al in the early 1990 [10] shows that ANN ability can be significantly improved through ensembling. The most accepted definition of artificial neural network ensemble is that ANN ensemble is a collection of a finite number of ANNs that are trained for the same task [11].

Over the last two decades, a lot of research works have been conducted for automated cancer diagnosis based on ANN. Zhi-Hua Zhou et al in [13] described an approach for utilizing the power of ANN ensembles in reliable applications such as diabetes, hepatitis and breast cancer.

Neural Network model for pattern recognition in medical diagnosis was described by Frenster, J.H. in [14]. G. Wilym et al in [15] developed an efficient neural network model for the diagnosis of carcinogenesis.

The benefit of artificial neural networks (ANNs) as decision making tools in the field of cancer was described in [16]. Chiou et al designed an ANN based system named HLND(hybrid lung cancer detection) to improve the accuracy of diagnosis and the speed of lung cancerous pulmonary radiology[17].

A system that employed an artificial neural network to detect suspicious regions in a low-resolution image was described in [18]. An automatic pathological diagnosis procedure named Neural Ensemble based Detection (NED) was implemented with an artificial neural network ensemble by Zhi-Hua Zhou et al in [19].

However in this work, nucleotides of the EGFR's deletion mutations were utilized to train and test different ANNs. These ANNs are ensembled to achieve an optimal informatics platform for the diagnosis of NSCLC. Section two of this paper explains the materials and methods we utilised in the research, section three discusses our experimental results while section four draws conclusions on the paper.

2. MATERIALS AND METHODS

The Tyrosine kinase(TK) domain is the region of the EGFR gene that is proned to mutation in NSCLC patients. The TK domain has 7 exons(exons 18-24), out of which exons 18-21 carry various somatic mutations in NSCLC patients[20].

The nucleotide ranges for exons 18 – 21 of the TK domain are shown in Table 1

Table 1. Nucleotides ranges in the EGFR TK domain

Exons	Amino acid ranges	Nucleotides ranges	Number of bases
18	687 – 728	2059 – 2184	126
19	729 – 761	2185 - 2283	99
20	762 – 823	2284 - 2469	186
21	824 – 875	2470 – 2625	156

Table 2 details the genomic profiles of NSCLC patients with microdeletion mutations that we extracted from[20]. The data in the table corroborate facts from other literatures on oncogenomics for the mutation patterns in NSCLC patients. From our data and from literatures, the two most common EGFR TK domain mutations are the in-frame deletion(2235-

2249del or E746-A750del) in exon 19 and the L858R(2573T<G) missense mutation in exon 21[20,21]. The structural patterns for the various deletion mutations' categories(Patient Category 1 to Patient Category 22) in Table 2 are illustrated in Figures 1,2 and 3. The deleted nucleotides are represented with (-).

Table 2. Microdeletion mutations

S/N	Patient Categories	Exons	Deleted Nucleotides	Number of Patients
1	1	19	c.2235-2249del	166
2	2	19	c.2236-2250del	60
3	3	19	c.2254-2277del	1
4	4	19	c.2240-2257del	33
5	5	19	c.2240-2254del	6
6	6	19	c.2239-2256del	7
7	7	19	c.2237-2251del	7
8	8	19	c.2238-2252del	2
9	9	19	c.2238-2255del	2
10	10	19	c.2237-2254del	3
11	11	19	c.2239-2247del	2
12	12	19	c.2239-2253del	1
13	13	19	c.2245-2253del	1
14	14	19	c.2253-2276del	2
15	15	20	c.2309-2310del	2
16	16	19	c.2236-2253del	1
17	17	18	c.2155-2156del	1
18	18	19	c.2238-2247del	1
19	19	19	c.2254-2255del	1
20	20	19	c.2235-2236del	2
21	21	19	c.2240-2251del	3
22	22	19	c.2229-2236del	1
			Total Patients	**305**

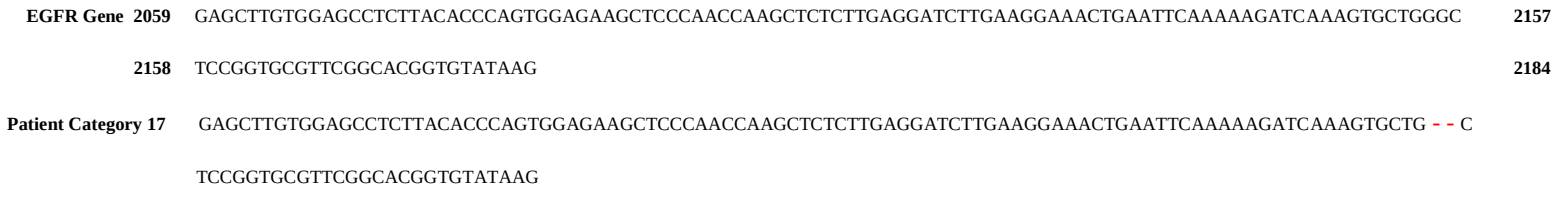
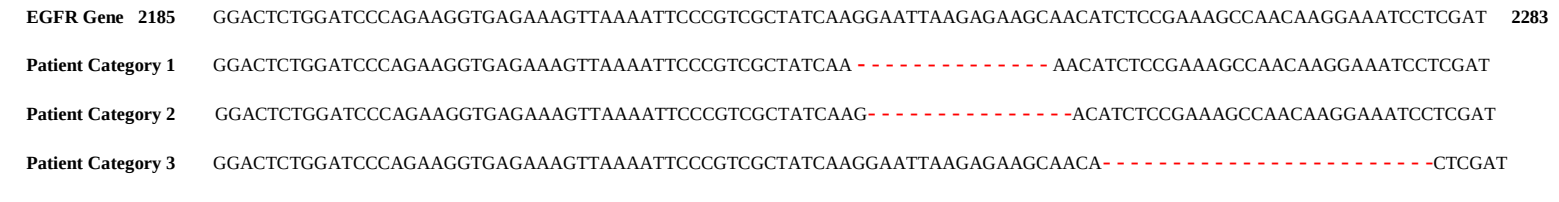


Fig 1: Exon 18 microdeletion mutation pattern(Patient Category 17)



Pateint Category 4 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGGAAT----------CGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 5 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGGAAT---------- CTCCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 6 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGGAA----------CCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 7 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGG---------- CATCTCCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 8 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGGA---------- ATCTCCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 9 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGGA----------TCCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 10 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGG---------- ------- CTCCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 11 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGGAA--------- GCAACATCTCCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 12 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGGAA---------- TCTCCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 13 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGGAATTAAGA--------- TCTCCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 14 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGGAATTAAGAGAAGCAAC----------CTCGAT

Patient Category 16 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAG---------- TCTCCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 18 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGGA---------- GCAACATCTCCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 19 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGGAATTAAGAGAAGCAACA -- TCCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 20 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAA -- AATTAAGAGAAGCAACATCTCCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 21 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAGGAAT ---------- CATCTCCGAAAGCCAACAAGGAAATCCTCGAT

Patient Category 22 GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGC --------- AATTAAGAGAAGCAACATCTCCGAAAGCCAACAAGGAAATCCTCGAT

Fig 2: Exon 19 microdeletion mutation patterns(Patient Categories 1,2,3,4,5,6,7,8,9,10,11,12,13,14,16,18,19,20,21,22)

EGFR Gene 2284 GAAGCCTACGTGATGGCCAGCGTGGACAACCCCCACGTGTGCCGCCTGCTGGGCATCTGCCTCACCTCCACCGTGCAGCTCATCACGCAGCTCATGCCCTT **2384**

2385 CGGCTGCCTCCTGGACTATGTCCGGGAACACAAAGACAATATTGGCTCCCAGTACCTGCTCAACTGGTGTGTGCAGATCGCAAAG **2469**

Patient Category 15 GAAGCCTACGTGATGGCCAGCGTGG -- AACCCCCACGTGTGCCGCCTGCTGGGCATCTGCCTCACCTCCACCGTGCAGCTCATCACGCAGCTCATGCCCTT

CGGCTGCCTCCTGGACTATGTCCGGGAACACAAAGACAATATTGGCTCCCAGTACCTGCTCAACTGGTGTGTGCAGATCGCAAAG

Fig 3: Exon 20 microdeletion mutation pattern(Patient Category 15)

We developed a novel algorithm for simulating each of the microdeletion mutations' category programatically and named it SimMicrodel. The algorithm was programmed in MATLAB and perfectly simulates mutations for the various microdeletion mutated patient categories. The primary purpose of this algorithm is to provide a basis for computer based emulation of any type of micro deletion mutations(once the exon and the deletion range are known) that may be introduced into the diagnostics system randomly and dynamically.

2.1 The Genomics Nucleotides Encoding and Normalization(GNEN) Algorithm

To comply with the mathematical structure of each ANN layer, input and output data is normally structured as a string or vector of numbers. One of the challenges in using ANNs is mapping how the real-world input/output(e.g. an image, a physical characteristics, a list of gene names, a prognosis) can be mapped to a numeric vector. This is what normalisation caters for. A novel Genomic Nucleotides Encoding and Normalization(GNEN) algorithm was

developed by us and programmed in MATLAB to encode the TK domain exons for normal and mutated genes. GNEN encodes each nucleotide code with their ASCII equivalent and then subsequently normalizes the values between -1 and 1. However, the deletionn mutation represented with(-) in the structural patterns are encoded as 0 (see Table 3).

Table 3. GNEN output for the four EGFR nucleotides and microdeletion mutations.

Nucleotide	ASCII Equivalent	Normalized Value
A	65	-1.0000
C	67	-0.7895
G	71	-0.3684
T	84	1.0000
deletion(-)	0	0

2.2 The Artificial Neural Network(ANN) Architecture for our Genomics Based ANNs

Figure 4 is the architectural network of our genomics based ANN. In figure 4, **p** is the input vector of R x 1 dimension where R is the number of rows.For batch processing, p is a matrix. **W** is the weigth matrix of dimension S x R where S is the number of neurons in the layer. **b** is the bias vector which is a weight with 1 as input. **n** is the weighted input into the transfer function. **a** is the layer output vector. **y** is the output vector from the network. The transfer function for the hidden layer is **tansig** and for the output layer is **purelin**. The network can be built with multiple hidden layers to enhance it's effectiveness.

Therefore the following ANNs(a-c) with the respective inputs and outputs were built and configured in MATLAB. Batch processing is adopted for the training.

a.) *Exon18MicroDelANN:* This has input matrix **x** of dimension 126 X 2. The elements range from $x_{1,1}$ to $x_{R,2}$ where R = 126. The number of columns in the matrix is based on the number of patients in the training data set.

b.) *Exon19MicroDelANN:* This has input matrix **x** of dimension 99 X 16. The elements range from $x_{1,1}$ to $x_{R,16}$ where R = 99. 16 represents the number of patients(both normal and mutated) in the training data set.

c.) Exon20MicroDelANN: This has input matrix **x** of dimension 186 X 2. The elements range from $x_{1,1}$ to $x_{R,2}$ where R = 186. 2 represents the number of patients(normal and mutated) in the training data set.

The output vector **y** has elements based on the number of patients in the training data set .

If y_i = 0 then the patient has non- mutated EFGR gene(normal).

If y_i = a (where a>0 and represent the number of deleted nucleotides divided by 10) the patient has microdeletion based mutated gene(NSCLC cancer).

2.3 The Microdeletion Mutations ANNs

Using the simulated and normalised genomic patterns with our novel algorithms (SimMicrodel and GNEN) and also the architecture in Figure 4, different ANNs were built based on each of the exons in EGFR's TK domain that are susceptible to mutation in our sample data(exons 18 – 20).

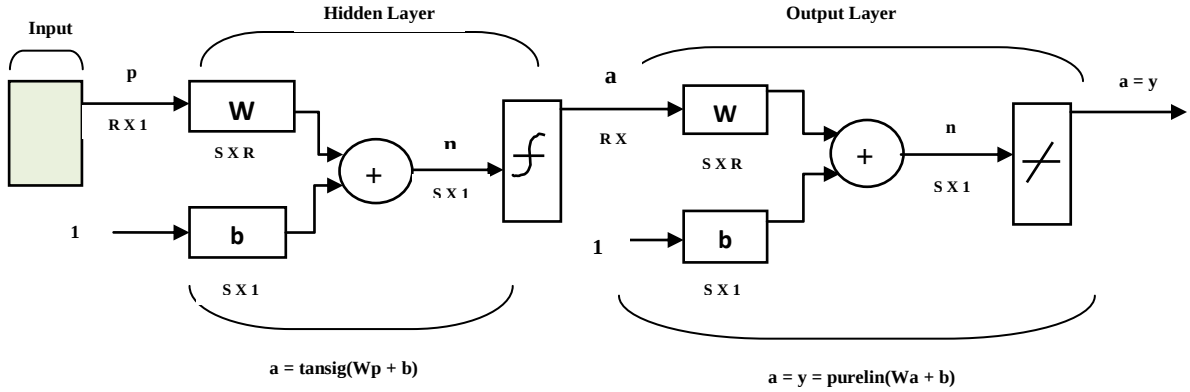


Fig 4: Architecture of the genomics based ANN(a feed forward artificial neural network)

3. RESULTS AND DISCUSION

The training datasets of the genomics based ANNs are shown in Table 4. The target vector element for each patient category was obtained by the number of deleted nucleotides divided by 10 .

Since the dataset is segmented into training and testing datasets. After proper training of the ANN with the training dataset and the appropriate configurations, the outcome of the various experiments performed by altering the number of hidden layers in the ANN configuration(Figure 4) and utilsing the training datasets (Table 4) is detailed in Table 5. Levenberg-Marquardt algorithm was used for all the experiments because our study in[22] shows that it is the best backpropagation training algorithm for the genomics based ANN.

After proper training,the various patient categories in the testing dataset were subjected to different configurations of the ANN for 5 different experiments. The outcomes of the tests are shown in Table 6. Table 7 shows the details of the errors and the minimum errors for each patient category.

Table 4. Training datasets target vector elements

S/N	Patient categories	Number of deletions	Target vector elements
1	Normal Patient	0	0
2	Patient Category 1	15	1.5
3	Patient Category 3	24	2.4
4	Patient Category 4	18	1.8
5	Patient Category 9	18	1.8
6	Patient Category 11	9	0.9
7	Patient Category 13	9	0.9
8	Patient Category 14	24	2.4
9	Patient Category 18	10	1
10	Patient Category 19	2	0.2
11	Patient Category 21	12	1.2
12	Patient Category 22	8	0.8

Table 5. Training experiments(Expts.1 to 5)

S/N	Training Parameters	Expt. 1	Expt.2	Expt. 3	Expt. 4	Expt.5
1	Epoch(iterations)	6	8	6	8	5
2	Time(seconds)	4	0	0	0	7
3	Performance(mse)	2.45×10^{-31}	0.00162	5.73e-19	1.92e-20	8.2e-26
4	Gradient	3.51×10^{-15}	2.45e-15	8.52e-16	2.44e-15	2.06e-12
5	Validation check	0	4	1	1	0
6	No of hidden layers	5	2	3	4	6

Table 6. Testing of the ANN with different patient categories testing datasets in experiments 1-5

	Normal Patient	Patient Category 2	Patient Category 5	Patient Category 6	Patient Category 7	Patient Category 8	Patient Category 10	Patient Category 12	Patient Category 16	Patient Category 20
No of deletions	0	15	15	18	15	15	18	15	18	2
Expected Output	0	1.5	1.5	1.8	1.5	1.5	1.8	1.5	1.8	0.2
Expt.1 Outputs	0	1.36	1.5	1.83	0.89	1.37	1.34	1.57	1.31	0.93
Expt.2 Outputs	0.0021	1.00	1.84	0.58	1.01	1.13	1.15	0.47	1.14	0.04
Expt.3 Outputs	6.16e-10	0.85	1.09	1.94	0.59	1.33	0.45	1.43	0.82	0.67
Expt.4 Outputs	5.18e-11	1.59	2.42	2.26	1.46	1.57	2.62	1.39	2.42	0.32
Expt.5 Outputs	5.61e-13	1.91	1.12	1.67	1.66	1.30	1.69	1.59	2.05	0.01

Table 7. Experiments 1-5 errors and minimum errors

Patient Categories	Expt.1 Errors	Expt.2 Errors	Expt.3 Errors	Expt.4 Errors	Expt.5 Errors	Min. Errors
Normal Patient	0	0.0021	6.16e-10	5.18e-11	5.61e-13	0
Patient Category 2	0.14	0.5	0.65	0.09	0.41	0.09
Patient Category 5	0	0.34	0.41	0.92	0.38	0
Patient Category 6	0.03	1.25	0.14	0.46	0.13	0.03
Patient Category 7	0.61	0.49	0.91	0.04	0.16	0.04
Patient Category 8	0.13	0.37	0.17	0.07	0.2	0.07
Patient Category 10	0.46	0.65	1.35	0.82	0.11	0.11
Patient Category 12	0.07	1.1	0.07	0.11	0.09	0.07
Patient Category 16	0.49	0.66	0.98	0.62	0.25	0.25
Patient Category 20	0.73	0.16	0.47	0.12	0.19	0.12

In Table 8, the appropriate genomics based ANN configurations(i.e.experiments) for determining the number of mutated(deleted) nucleotides for each patient category are illustrated.This is based on the ANN configuration(s) that produced the minimum error for the patient category. The training outputs for experiment 5 in MATLAB R2008a are shown in Figures 5,6 and 7.

Table 8. Patient categories and the appropriate ANN configurations(experiments)

S/N	Patient Categories	Appropriate Experiments (ANN Configurations)
1	Normal Patient	Experiment 1
2	Patient Category 2	Experiment 4
3	Patient Category 5	Experiment 1
4	Patient Category 6	Experiment 1
5	Patient Category 7	Experiment 4
6	Patient Category 8	Experiment 4
7	Patient Category 10	Experiment 5
8	Patient Category 12	Experiment 1 or 3
9	Patient Category 16	Experiment 5
10	Patient Category 20	Experiment 4

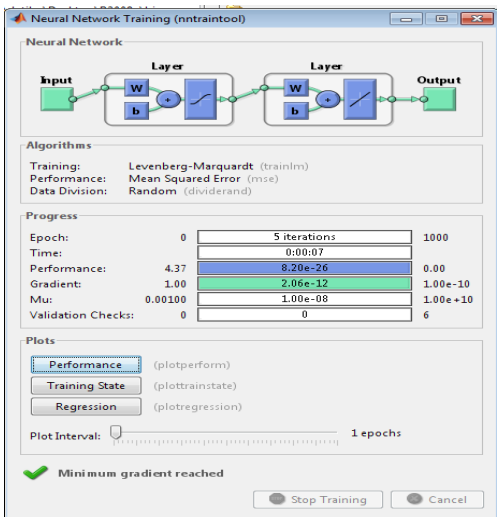


Fig5: ANN training output for experiment 5 in MATLAB R2008a

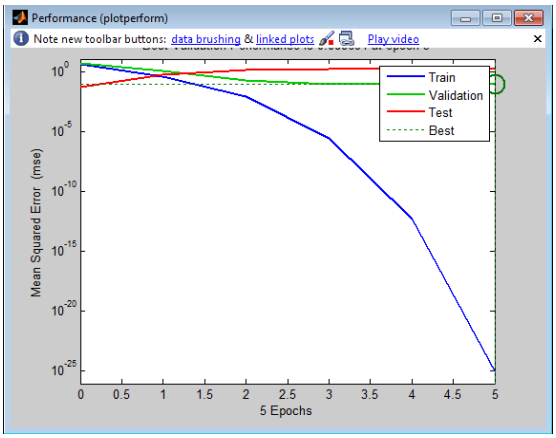


Fig 6: Performance plot (mse versus epochs) for experiment 5

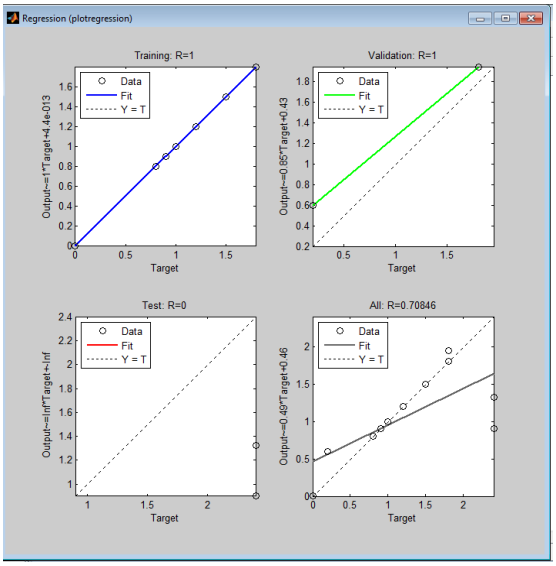


Fig 7: Regression plot for experiment 5

4. CONCLUSION

Tables 7 and 8 show the results of the different genomics based ANNs. Our results show that ensembling the ANNs and utilising the one with the minimum error to predict the number of deleted nucleotides in the cancerous patient is highly optimal. We therefore tag this ANN ensembling approach as minimum error voting ensembling. Further work on this research will increase the number of experiments(ANN configurations) so as to obtain predictions that achieve almost zero error. This will further enhance the precision of our genomics based ANN for diagnosis of non-small cell lung cancer(NSCLC) electronically.

5. REFERENCES

[1] Genetic Disease Information Pronto, www.ornl.gov/sci/techresources/Human_Genome/medicine/assist.html

[2] Genetics of Cancer, www.britannica.com

[3] Rogerio C. L., Winfield A. B., Carolyn M., Progress in the treatment of lung cancer. CANCER Care Help and Hope.

[4] Breathnach O.S., Freidlin B., Conley B. 2001. Twenty-two years of phase III trials for patients with advanced non-small-cell lung cancer: Sobering results. Journal of Clinical Oncology 19, 1734-1742.

[5] Lynch T.J., Bell D.W., Sordella R. 2004. Activating mutations in the epidermal growth factorreceptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N. Engl J. Med 350, 2129-39.

[6] Paez J.G., Janne P.A., Lee J.C., 2004. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304 1497-500.

[7] Sharma S. V. 2007. Epidermal growth factor receptor mutations in lung cancer. Nature Rev. Cancer 7, 169–181.

[8] Eunice L. K., Janusz J., Sarah P. T. Epidermal Growth Factor Receptor Kinase Domain Mutations in Esophageal and Pancreatic Adenocarcinomas, www.aacrjournals.org

[9] Pao W., Miller V.A. 2005. Epidermal growth factor receptor mutations, small-molecule kinase inhibitors, and non-small-cell lung cancer: current knowledge and future directions. Journal of Clinical Oncology, vol. 23, no.11, 2556-68.

[10] Hansen L.K., Salamon P. 1990. Neural network ensembles. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 10, 993-1001.

[11] Sollich P., Krog A. 1996. Learning with ensembles. In Advances in Neural Information Processing Systems 8, Cambridge, MA,MIT Press 190-196.

[12] Naguib R.N.G., Sherbet G.V. 1997. Artificial Neural Networks in Cancer Research. Pathobiology vol. 65, no. 3, 129-139.

[13] Zhi-Hua Zhou, Yuan Jiand 1990. Medical diagnosis with C4.5 rule preceded by ANN ensemble, Pattern Analysis and Machine Intelligence, IEEE Transactions, vol.12, no.10, 993 - 1001.

[14] Frenster, J.H. 1990. Neural Networks for Pattern Recognition in Medical Diagnosis. Annual International Conference in the IEEE Engineering in Medicine and Biology Society, vol. 12, no.3, 1423-1424.

[15] Naguib R.N.G., Robinson M.C., Neal D.E., Hamdy F.C. 1998. Neural network analysis of combined conventional and experimental prognostic markers in prostate cancer:a pilot study. British Joumal of Cancer vol. 78, no. 2, 246-250

[16] Paulo J. L., Azzam F. G. 2006. The use of artificial neural networks in decision support in cancer;A systematic review.Neural Networks vol. 19, no.4. .

[17] Chiou Y.S.P., Lure Y.M.F., Ligomenides P.A. 1993. Neural network image analysis and classification in hybrid lung nodule detection (HLND) system, Proceedings of the IEEE-SP Workshop on Neural Networks for Signal Processing ,517-526.

[18] Penedo M.G., Carreira M.J., Mosquera A., Cabello D. 1998. Computer-aided diagnosis: a neural-network-based approach to lung nodule detection. IEEE Trans. Medical Imaging vol. 17, no. 6, 872-880.

[19] Zhi-Hua Zhou, Yuan Jiang, Yu-Bin Yang, Shi-Fu Chen. Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles. National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, P.R.China.

[20] Pao W., Miller V., Zakowski M., Doherty J., 2004. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc. Natl. Acad. Sci. U S A. vol. 101, no. 36 13306-13311.

[21] MDL EGFR Mutation Database. City of Hope, www.EGFR.com/EGFRMutationDataByMutAug2005.pdf

[22] E. Adetiba, J.C. Ekeh, V.O. Matthews, S.A. Daramola, M.E.U Eleanya 2011. Estimating an Optimal Backpropagation Algorithm for Training an ANN with the EGFR Exon 19 Nucleotide Sequence: An Electronic Diagnostic Basis for Non-Small Cell Lung Cancer(NSCLC). Journal of Emerging Trends in Engineering and Applied Sciences, vol. 2, no.1, 74-78.