

Mining Method for Cancer and Pre-Cancer Detection Caused by Mutant Codon 248 in TP53**Deeman Yousif Mahmood¹, Prof. Dr. Ayad Ghany Ismaeel², Prof. Dr. Abbas Hassan Taqi³**¹ Department of Computer Science, University of Kirkuk² Department of Computer Technical Engineering, Al-Kitab University³ Department of Mathematics, University of Kirkuk**Article Info**

Received Dec8, 2018

Keyword:Mining
Machine Learning
Quick Propagation
TP53
Cancer Detection**ABSTRACT**

Process of prediction has a substantial function in detecting and efficient protection of cancer. The tumor suppressor P53 is approximately near 50% of all human beings tumors due to the mutations which is appear in the TP53 gene to the cells within updated UMD TP53 Mutation Database Oct. 2017 [1], it is so difficult working with prime data (in excel) to predict and diagnosis cancers. In this research a functional model of mining approach and Artificial Neural Network which is proposed to predict cancer and pre-cancer caused by specific codon mutation (each codon has hundreds mutations cause cancers) of tumor protein P53, and applied this approach on mutability of hotspot codon 248 (exon 7), CGG. The Quick Propagation mechanism has been used for training and testing the Neural Network structure to determine the accuracy of the proposed architecture. This research procedure demonstrates that Neural Network based prediction of Cancer and Premalignant Disease (pre-cancer) of mutated codon 248 and manifests perfect performance in the prognosis of the mutation situation to pre-cancer or cancer in general. Using of data mining preprocessing steps and pattern extraction to construct the prediction model by selecting (8) out of (132) new TP53 gene database fields in order to classify the cases to the target class pathology (Cancer, Pre-cancer) using these fields. A high professional Neural Network software simulation (Alyuda NeuroIntellegence) is used to build the classifier and Neural Network, the testing and experimental results from the proposed architecture shows that using Quick Propagation algorithm is very accurate in term of accuracy and minimum error rates showing the results of accuracy (99.97%, 100%, 99.85%) for (Train, Validation and Test) phases respectively with error rate of (0.0003, 0, 0.0015) for (Train, Validation and Test) phases respectively.

Corresponding Author:Deeman Yousif Mahmood,
Departement of Computer Science,
University of Kirkuk,
Sayada, 36001, Iraq - Kirkuk.
Email: demo_y86@yahoo.com

1. Introduction

Cancer disease considered as a standout amongst the vastly known and complicated infections now a days, since it takes place due to various organic and physical activities and responses. Computer based paradigms of disease detection are being produced to help in both clinical recipes and biological originality [2]. The p53 gene holds mutations in about 50–60% from humanbeing cancers and around 90% of the mutations encrypt missense mutated proteins which extend about 190 diverse codons located in the DNA-binding range for both of gene and protein. These alterations results in a protein that has a miniature abiliasty to link to a particular DNA sequence which is organizing the p53 transcription route [3]. The epidermal tumour oppressor gene p53 situated on the short arm of chromosome 17(17p13), further minutely, TP53 gene is located in the range of base pair 7,571,719 to base pair 7,590,867 on the chromosome 17 as shown in Fig. 1, and since its invention in 1979, the p53 became the topic of keen scrutiny, and yet such is the intricate attitude of this protein, it sticks to defy researchers in spite of roughly four decades of explore and research [4].

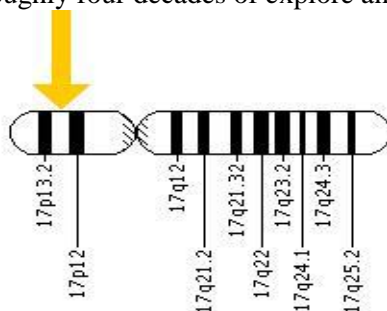


Figure 1. TP53 gene position on the short arm of Chromosome 17

The p53 tumor suppressor/transcription operator controls enormous variety of cellular functions and vigors, such as cell cycle, apoptosis, DNA reform and metabolism" [5]. The entire length variant/isoform generates a protein of 393 amino acids (Codon), and plurality of TP53 mutations alter monocular amino acids which is considered as protein building blocks in TP53, that results in the procreation of a modified protein version which is not able to link efficiently to DNA and the defective protein will be able to construct up at the kernel of the cells and prohibit them from subjecting apoptosis as a reaction to DNA spoiling, while spoiled cells carry on growing with dividing on an un streamline direction, that may results in cancerous tumors [4].

Codon 248 which is located on exon 7 in p53 gene sequence as shown in Fig. 2 depending on the International Agency for research on Cancer (IARC), and acts as a hotspot mutant codon that affect normal protein activities and functions and contribute to tumor malignancy and chemoresistance [6].

12601	cctctgattc	ctcactgatt	gctcttaggt	ctggcccctc	ctcagcatct	tatccgagtg	<u>exon 6</u>
12661	gaaggaaatt	tgctgtgga	gtatttggat	gacagaaaca	cttttcgaca	tagtgggtg	
12721	gtgcccattg	agccgcctga	ggtctggttt	gcaactgggg	tctctgggag	gaggggttaa	
12781	gggtggttgt	cagtgccctt	ccaggtgagc	agtagggggg	ctttctctct	ctgcttattt	
12841	gacctcccta	taaccccatg	agatgtgcaa	agtaaatggg	ttaactattt	gcacagttag	
12901	aaaaactgaa	gcttacagag	gctaagggcc	tcccctgctt	ggctggggcg	agtggtctcat	
12961	gcctgtaatc	ccagcaactt	gggaggccaa	ggcagggcga	tcacgaggtt	gggagatcga	
13021	gaccatcctg	gctaaccgtg	aaaccccgtc	tctactgaaa	atacaaaaaa	aaaattagcc	
13081	gggctggtg	ctggccacct	gtagtccacg	ctactcggga	ggctgagaaa	ggagaatggc	
13141	gtgaacctgg	gcggtggagc	ttgcagttag	ctgagatcac	gccactgcac	tccagcctgg	
13201	gcgacagagc	gagattccat	ctcaaaaaaa	aaaaaaaag	gcctcccctg	cttgccacag	
13261	gtctcccaaa	ggcgcaactg	cctcatcttg	ggcctgtggt	atctcctagg	<u>ttggctctga</u>	<u>exon 7</u>
13321	<u>ctgtaccacc</u>	<u>atccactaca</u>	<u>actacatgtg</u>	<u>taacagttcc</u>	<u>tgcattgggg</u>	<u>gcattgaaccg</u>	
13381	<u>gaggccacc</u>	<u>ctcaccatca</u>	<u>tcacactgga</u>	<u>agactccagg</u>	<u>tcaggagcca</u>	<u>cttgccaccc</u>	
13441	tgcaactggt	cctgctgtgc	cccagcctct	gcttgctctt	gaccctggg	cccacctctt	
13501	accgatttct	tccataactac	taccatcca	cctctcatca	catccccggc	ggggaatctc	
13561	cttactgctc	ccactcaagt	ttcttttctc	tggtttggg	acctcttaac	ctgtggcttc	
13621	tcctccacct	acctggagct	ggagcttagg	ctccagaaa	gacaagggtg	gttgggagta	
13681	gatggagcct	ggttttttaa	atgggacagg	taggacctga	tttcttact	gcctcttget	
13741	tctcttttcc	tatcctgagt	<u>agtggtaatc</u>	<u>tactgggacg</u>	<u>gaacagcttt</u>	<u>gaggtgctg</u>	<u>exon 8</u>
13801	<u>tttgtgcctg</u>	<u>tcctgggaga</u>	<u>gaccggcgca</u>	<u>cagaggaaga</u>	<u>gaatctccgc</u>	<u>aagaaagggg</u>	
13861	<u>agcctcacca</u>	<u>cgagctgccc</u>	<u>ccagggagca</u>	<u>ctaagcgagg</u>	<u>taagcaagca</u>	<u>ggacaagaag</u>	
13921	cggtggagga	gaccaagggt	gcagttatgc	ctcagatcca	cttttatcac	ctttctctgc	
13981	ctcttctcta	<u>gcactgcccc</u>	<u>acaacaccag</u>	<u>ctcctctccc</u>	<u>cagccaaaga</u>	<u>agaaccact</u>	<u>exon 9</u>
14041	<u>ggatggagaa</u>	<u>tatttcaacc</u>	<u>ttcaggtact</u>	<u>aagtcttggg</u>	<u>acctcttacc</u>	<u>aagtggaaa</u>	

Figure 2. Position of Codon 248 on exon 7 in P53 gene sequence

2. Related Works

Ayad G. Ismaeel, with Raghad Z. Yousif in (2015), suggested a mechanism for classifying, diagnosing patient mutations with predicting mutations location of the diseased person. The TP53 gene databases were employed with (6) selected columns of the UMD_Cell_line_2010 database, for training and testing Quick Propagation Network (QPN) and the mining technique was based on training the Quick Propagation, that is a refinement of the traditional back propagation network, by using the number of nodes (283-141-1) that used as an (input - hidden – output) layers respectively. Following results were obtained from the learning phase for (train, test, and validation set): R-squared = (0.9987), Correlation = (0.9993), with Mean of Absolute Relative Error = (0.0057) [7].

Zahraa N. Shahweli, Ban. N. Dhannoon, Rehab S. Ramadhan in (2017), proposed a silico molecular classifying approach for breast and prostate cancers by the mean of Back Propagation Network, by adopting seven datasets to assess the proposed model, five datasets from the UMD TP53 database and two sets of the (IARC TP53 database), the back propagation using hybrid pattern with 5 fold cross validation and sets of validation used for prediction and classification of breast and prostate cancers of patients using molecular mutations situated in the TP53 gene. The performance evaluation of the proposed model for both UMD TP53 and IARC TP53 respectively with accuracy = (98 , 96.7), specificity = (96.6 , 97.3) , and sensitivity (Recall) = (97 , 96.5) [2].

Zahraa N. Shahweli, Ban. N. Dhannoon (2017), proposed an enhanced method of the Relief algorithm named ReliefK and output favorable results for feature selection. Back Propagation Neural Network (BPNN) also used with this method to predict and classify cancer based on the mutations that were listed in the somatic and germline mutations of IARC TP53 database. Five measures, including accuracy (Acc), sensitivity (Sn), F-measure, Matthew correlation coefficient (MCC), and specificity (Sp), have been used to assess the proposed algorithm. The proposed method of feature selection ReliefK and BPNN results MCC of 1 and 0.88 for IARC TP53 somatic and germline mutations, respectively [8].

Chaurasia and Pal (2014), analyzed the performances of several data mining techniques. The breast cancer data classification can be utilized for finding the result of several disease or disclosing the common nature of cancer disease. Several data mining algorithms were used for analyzing cancer disease, the present novel approach used to find out the compare performances of decision tree classifier such as K-Nearest Neighbor Classifier, and Sequential Minimal Optimization (SMO), and Best First Tree. The result shows that the performance of SMO provides good result compare to other classifier in terms of accuracy, low error rate and performance [9].

3. Dataset and PreProcessing

There are sundry primitive databases such as protein dataset and excel genome, TP53 gene and its tumor protein with its mutations that cause diseases (cancers) is one of these types of dataset. In this study, modern bio-database (last updating at Oct. 2017) of TP53 (tumor protein P53) has been used which contains a big prime data of mutations (Excel file form). This bio-database contains about 80,400 tumors (mutated) records in different codons (in file called UMD_mutations_US) within “UMD TP53 Mutation Database” [1], [10]. The P53 new dataset consist of 132 fields and due to irrelevant fields and data quality assurance selection and preprocessing steps of data mining is required as follow:

1- Data Selection

Since the study focuses on Codon 248 mutations, the target data will be all records with this domain which is about (5120) records from the whole dataset, and in addition to data selection this step consider as dataset reduction on records (Row) level.

2- Data Cleaning

This step includes eliminating missing values from the new selected dataset named (Codon 248 dataset) and removing noise and inconsistent data.

3- Data Normalization

Data normalization is an essential part of knowledge data discovery for changing the data scale and standardized the format and for the raw data used in this research Z-score normalization has been implemented since the Z-score is an effective method of normalization for homogeneous data belonging to same range (i.e. in our case codon 248), all the data parameters related to the codon 248. This method that gives the normalized range of data from the original unorganized data using the measures of central tendency like "mean and standard deviation" then the Parameter is called as Z-score Normalization. So the unstructured data can be normalized using z-score parameter, as per given in the following equation [11]:

$$x_{ij} = \frac{x_{ij}^* - m_j}{s_j} \quad (1)$$

The form of normalization that is based on the z-score formula, scales and translates the feature x_{ij} , where m_j is the sample mean and s_j is the standard deviation of the attribute j respectively.

4- Feature Subset Selection

Feature selection method is utilized for distinguishing the input fields that are not profitable and do not have any significant contribution to the performance of the System [12]. The UMD TP53 Mutation Database consist of 132 features, and most of these features are irrelevant to the prediction function and time consuming if all features considered for computation. The abstraction of trivial fields will progress the Neural Network evaluation metrics. The Backward stepwise feature selection method is used for determining the importance of the input features, which starts with all features and eliminating one input feature at each step by finding a feature with least deterioration the network fulfillment to be eliminated from the input features set [12], and presented in Fig. 3.

Features that are nominated to be the input set for the model based on their importance are shown in table 1.

Table 1. Input Set features

Column No.	Feature Name	Description
17	Start_Cdna	Mutation start coordinate using the p53 cDNA as reference (location 1 points to the A of the start of ATG)
18	End_cDNA	Mutation end coordinate using the p53 cDNA as reference (location 1 points to the A of the start of ATG)
21	Genome_base_coding	Nucleotide at the start position of the mutation.
22	cDNA_variant	Mutation nomenclature according to HGVS standards using the coding sequence as reference (location 1 points to the A of the start of ATG)
26	Mutant_AA_1	Mutant amino acid: 1-letter nomenclature
27	Mutant_AA_3	Mutant amino acid: 3-letter nomenclature
36	Mutant_codon	<u>NNN</u> : Sequence of the mutated codon.
		Del: exonic deletion
		Ins: exonic insertion
		Indel: complex event that involves an exonic insertion and a deletion.
		Splice: mutation that aims the essential address of AG splice acceptor site or GT splice donor site.
		NR: not relevant, mutations targeting intronic sequence, <u>5'UTR or 3'UTR</u> .
49	Sample_origin	Adjacent tissue, Adjacent tissue (stroma), Cell line, Circulating tumour cells, Extra cellular DNA, Normal tissue, Pathological tissue, Peripheral blood lymphocytes, Tumour, Xenograft

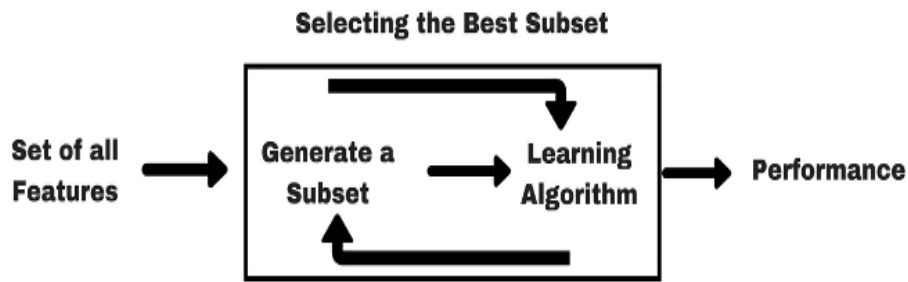


Figure 3. Backward stepwise feature selection

4. Proposed Mining Method of Cancer and Pre-Cancer Caused by Mutant Codon 248 of TP53

The field of artificial neural networks has been exceedingly elaborated and has been extensively used in many applications of artificial intelligence and because of their outstanding ability of "self learning and adapting", they get much concern from the scientific community, and now a days they have been considered as a pivotal components of many systems and are deemed as a robust tool for the solution of various type of problems [13], [14]. Two major types of learning process are exist namely supervised learning and unsupervised learning. The neural network in the supervised learning know the output of the learning function (target) and the fitting of weight coefficients is done in such a manner that the desired and the computed outputs are tend to be as close as possible [15].

This paper is concerned with supervised learning and Quick Propagation Neural Network has been implemented as a mining method for training and testing of Cancer and Pre-Cancer diagnosis using the input selected features with the target output. This case study supposed that there is a mutation in Codon 248, the rule of the neural network is to classify the mutation into Cancer or Pre-Cancer via the Quick Propagation algorithm. The quick propagation is Newton’s method based to determine the solution and this algorithm is an enhancement of back propagation algorithm. The learning criteria of the Quick Propagation depends of picking out a distinction of this link coefficient by combining two traditional approaches used in [16]:

1. Learning rate regulation based on the historic improvement of the learning rate up to the effective value for each weight either globally or apart.
2. Second derivative calculation of the error for the sake of each weight.

All the weights are updated separately in the quick back propagation algorithm as used in [16], [17]. The resulting weight updating rule which is:

$$\begin{cases} \Delta w_{ij}^t = \Delta w_{ij}^{t-1} \frac{\nabla L^t}{\nabla L^{t-1} - \nabla L^t} & (2) \\ w_{ij}^t = w_{ij}^{t-1} + \alpha \Delta w_{ij}^t & (3) \end{cases}$$

As L is the function of total loss "cost function", while the gradient have to calculate in respect to the connection between neurons (i belongs to higher layer and j belongs to lower layer), while learning rate represented by "alpha".

The Quick propagation implemented via simple neural network in a simplest way as shown in Fig. 4 the hypothesis of the neural network:

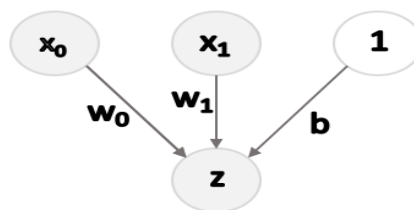


Figure 4. Simple Neural Network

The bias b similar to any other weight can be learned, where the total loss function based on the "cross-entropy" is:

$$L = - \sum_i y_i \log \tilde{y}_i + (1 - y_i) \log (1 - \tilde{y}_i) =$$

$$= - \sum_i y_i \log \sigma(z_i) + (1 - y_i) \log (1 - \sigma(z_i)) \tag{4}$$

$$\text{where } z_i = \bar{w}^T \cdot \bar{x}_i \text{ and } \sigma(z_i) = \frac{1}{1 + e^{-z_i}} \tag{5}$$

Where the generic gradient calculated as follow:

$$\nabla L_i = \sum_i (y_i - \sigma(z_i)) x_i \tag{6}$$

5. Results and Performance Measures

The Dataset have been analyzed in this research using Alyuda Neurointelligence tool [18]. The dataset size fixed at 5095 records out of 5120 records, after eliminating records of missing values. Through the data analysis, Pathology column is set to be the target, other eight columns will be input columns. The dataset is divided in to subsets represented by training, validation and test set as the following percentages:

- 1- Training Set = 60%
- 2- Validation Set = 20%
- 3- Test Set = 20%

The dataset errors presented in Fig. 5:

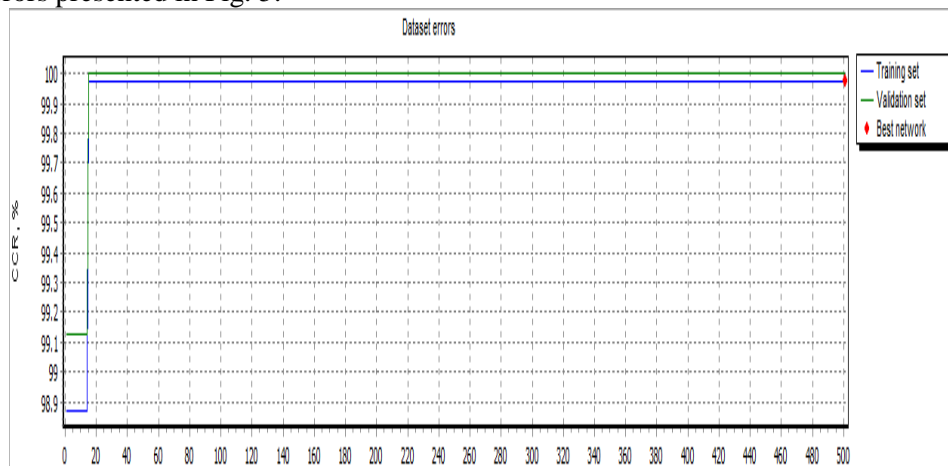


Figure 5. Dataset Errors

At the Training, Validation, and Test stages the performance of the model will be estimated. The performance of machine learning model evaluation indices as follows:

- 1- Accuracy.
- 2- Precision.
- 3- Recall.
- 4- True Positive (TP): correctly identified.
- 5- True Negative (TN): correctly rejected.
- 6- False Positive (FP): incorrectly identified.

Precision and Recall factors are independent from the training and testing sample size. Derivation of these metrics are from the confusion matrix which is a standard data structure [19]. A simple confusion matrix of two classes Premalignant Disease (Pre-Cancer), Cancer classification of our case is represented as in table 2.

Table 2. Confusion Matrix

Actual Class	Predicted Class		
	Activity	Premalignant Disease	Cancer
Premalignant Disease	TP		FN
Cancer	FP		TN

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

Results are obtained for each step of the learning (training), validation and testing phases are as follow:

1- Learning (Training) Phase:

Learning phase done on three ways using the same dataset to train the network and construct the model using three different input activation function which are:

- i. Logistic
- ii. Hyperbolic Tangent
- iii. Linear

In this step which is about learning the Quick Propagation neural network using the training set that is consist of 60% of the whole data by the mean of above input activation functions results the same prediction rates for the all three function since the sample dataset is well normalized and all parameters belonging to the same codon 248 (i.e. homogeneous data) and the result was:

Activity	Premalignant Disease	Cancer
Premalignant Disease	41	1
Cancer	0	3677

From the obtained values of confusion matrix:

Accuracy = 0.9997 = 99.97%

Error Rate = 1 – Accuracy = 1 – 0.9997 = 0.0003

Precision = 100%

Recall = 97.62%

The ROC Curve of Training phase presented in Fig. 6:

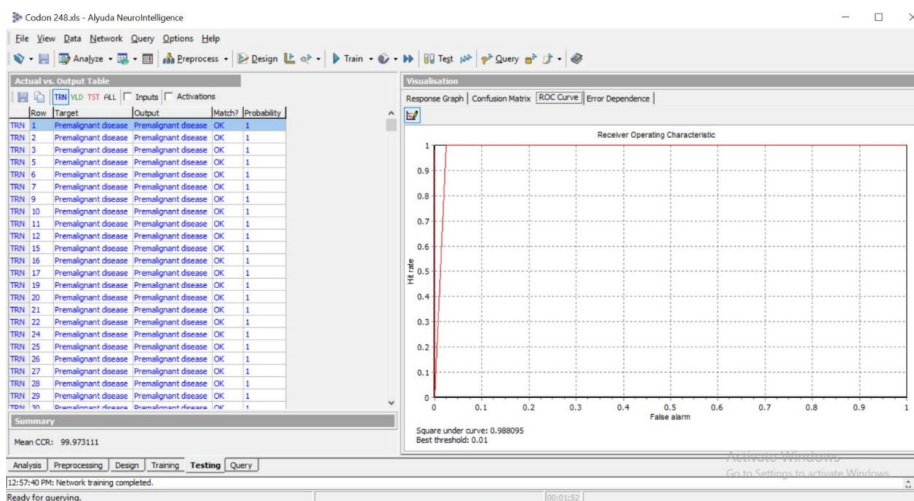


Figure 6. ROC Curve of Training phase

2- Validation Phase

The results of this step which is about validating the model and tune the parameters and this set consist of 20% of the whole data was:

Activity	Premalignant Disease	Cancer
Premalignant Disease	6	0
Cancer	0	682

From the obtained values of confusion matrix:

Accuracy = 100%

Error Rate = 0%

Precision = 100%

Recall = 100%

The ROC Curve of the validation phase presented in Fig. 7:

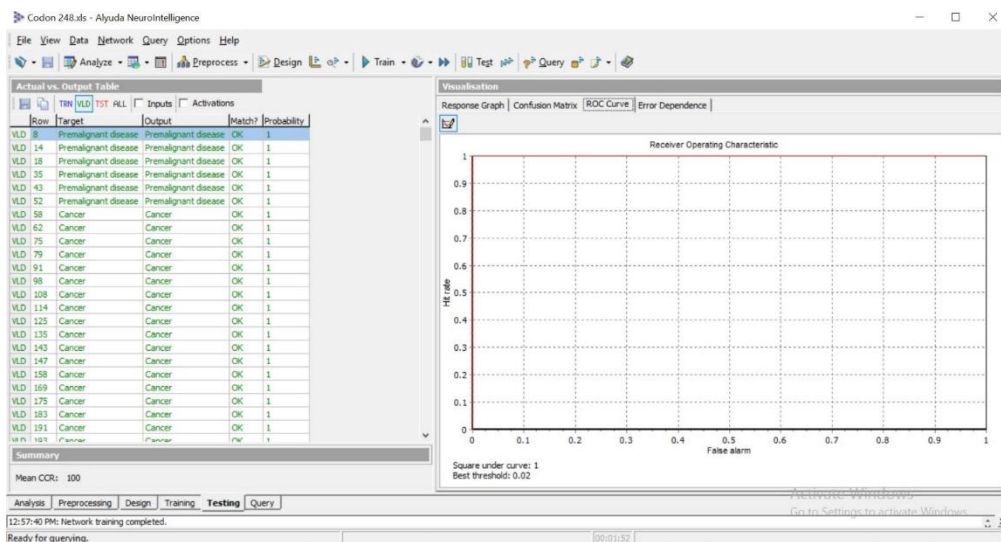


Figure 7. ROC Curve of the validation phase

3- Testing Phase

The results of this step which is about testing the Quick Propagation neural network after learning and validating the model by the mean of Testing set that is consist of 20% of the whole data was:

Activity	Premalignant Disease	Cancer
Premalignant Disease	5	1
Cancer	0	682

From the obtained values of confusion matrix:

Accuracy = 0.9985 = 99.85%

Error Rate = 1 – Accuracy = 1 – 0.9985 = 0.0015

Precision = 100%

Recall = 83.33%

The ROC Curve of Testing phase presented in Fig. 8:

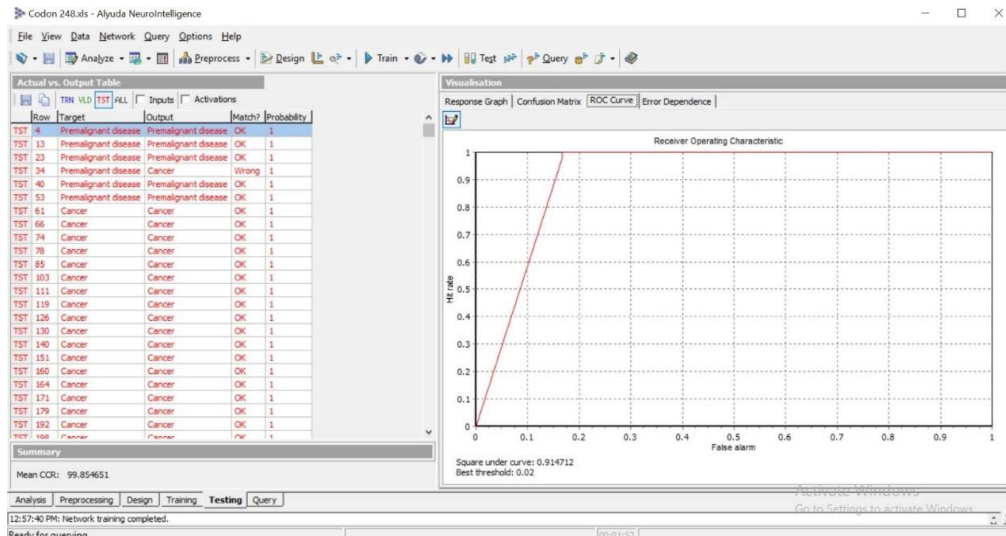


Figure 8. ROC Curve of Testing phase

According to the testing algorithm used in this research which is the QPN to predict the type of mutation in codon 248 of P53 whether it is Cancer or Pre-Cancer after training and validating the model. The proposed mining algorithm allows to predict the situation of the mutations by filling each field parameter manually as shown on Fig. 9.

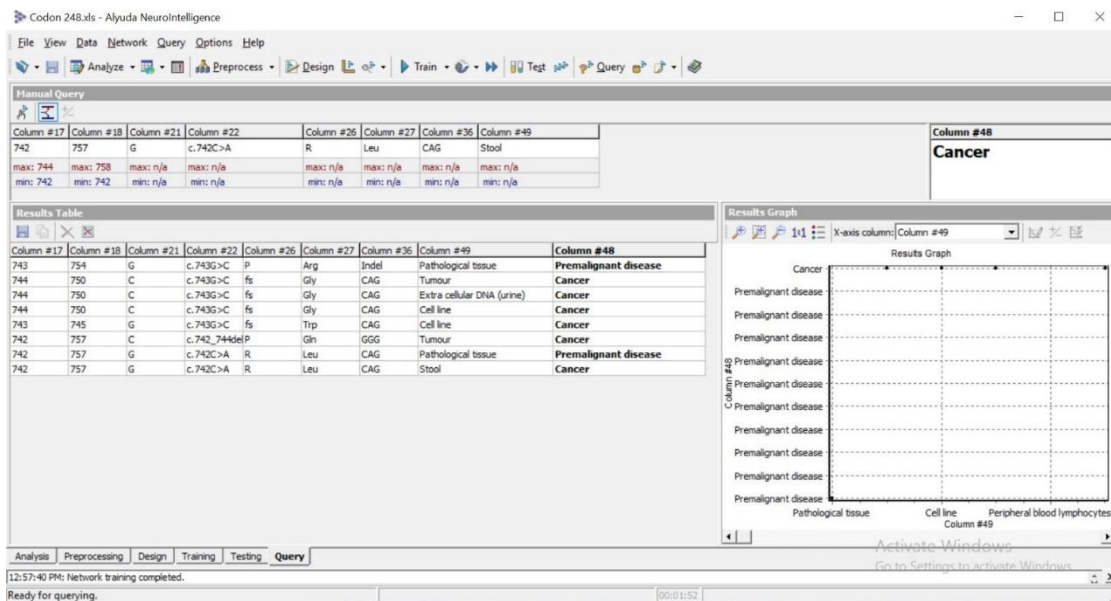


Figure 9. Shows the prediction model of codon 248 mutations

6. Discussion the Results

Comparative of the results of our model via the proposed mining method with other related methods (in case of same technique comparing with QPN, while in case of different techniques comparing with two other techniques of Standard Back Propagation Network in term of predicting cancer caused by mutated P53, via this comparison will explain the suggested method was effective in diagnosis cancers comparing to other techniques as shown in table 3:

Table 3. Comparison of mining method with other techniques

Proposed Model	Ayad and Raghad (2015) [7]	Zahraa, et. al. (2017) [2]	Zahraa and Ban (2017) [8]
Using QPN with a topology of: (75-37-1)	Using QPN with a topology of: (283-141-1)	Using Standard Back Propagation Network	Using Standard Back Propagation Network
Using Modern Biodata of TP53: UMD_mutations_US	Using old version of UMD TP53 mutation database	Using old version of UMD TP53 and IARC TP53 databases	Using IARC TP53 databases
Work on specific codon of TP53 (Codon 248)	Work on all TP53 codons	Work on all TP53 codons	Work on all TP53 codons
Using only records related to codon 248 from the whole new dataset	Using the whole dataset	Using the whole dataset	Using the whole dataset
Prediction Cancer and Pre-Cancer based on the mutations of Codon 248	Diagnose and predict cancers via mutation in TP53 sequence	classify breast and prostate cancers of patients using molecular mutations situated in the TP53 gene	classify cancer based on the mutations registered (IARC) TP53 database

7. Conclusion

The proposed mining method of prediction cancers and pre-cancer by mutations in codon 248 of TP53 demonstrates following conclusions:

- 1- The Quick Propagation Neural Network has been implemented for classification of mutant codon 248 and its high accuracy for classifying mutations in to Cancer and Premalignant Disease.
- 2- The mining method is accurate in predict cancers and pre-cancer via codon 248 mutations, by training and testing the quick propagation with minimal number of inputs: (8) features of a huge number of fields total (132) field of UMD_mutations_US TP53 dataset.
- 3- The objective of this work is to mine in a big data set such as the modern biodata of TP53 (tumor protein P53) named UMD_mutations_US and the focal point of the research is concerned about hotspot mutant codon 248 and detection of the infection caused by this mutant codon either cancer or pre-cancer.
- 4- Separating Codon 248 from the whole data set will enhance the learning and testing of the constructed model and through the experiments of this work and the corresponding results obtained from the model it could emphasized that the updated P53 data set is a rich biodata for mining purposes and can help studying each codon (amino acid) alone to determine the type of cancers that could be caused by TP53.

References

- [1] <http://p53.fr/tp53-database/mutation-database>, "UMD TP53 Mutation Database" Oct, 2017.
- [2] Zahraa N. Shahweli, Ban N. Dhannoon, Rehab S. Ramadhan (2017), "In Silico Molecular Classification of Breast and Prostate Cancers using Back Propagation Neural Network", *Cancer Biology*, 7(3).
- [3] Evan H. Baugh, Hua Ke, Arnold J. Levine, Richard A. Bonneau and Chang S. Chan (2018), "Why are there hotspot mutations in the TP53 gene in human cancers", *Cell Death and Differentiation* 25, 154–160.
- [4] Ayad Ghany Ismaeel (2013), "New Approach for Prediction Pre-cancer via Detecting Mutated in Tumor Protein P53", *International Journal of Scientific & Engineering Research*, Volume 4, Issue 10.
- [5] Jiandong Chen (2016), "The Cell-Cycle Arrest and Apoptotic Functions of p53 in Tumor Initiation and Progression", *Cold Spring Harb Perspect Med* 2016.
- [6] Praveen K. Jaiswal, Apul Goel, Rama D. Mittal (2011), "Association of p53 codon 248 (exon7) with urinary bladder cancer risk in the North Indian population", *BioScience Trends*, 5(5):205-210.

- [7] Ayad. Ghany Ismaeel, Raghad. Zuhair Yousif (2015), "Novel Mining of Cancer via Mutation in Tumor Protein P53 using Quick Propagation Network", International Journal of Computer Science and Electronics Engineering (IJCSEE) Volume 3, Issue 2.
- [8] Zahraa N. Shahweli, Ban N. Dhannoon (2017), "NEURAL NETWORK WITH NEW RELIEF FEATURE SELECTION FOR PREDICTING BREAST CANCER BASED ON TP53 MUTATION", International Research Journal of Computer Science (IRJCS), Issue 12, Volume 4.
- [9] Chaurasia V., Pal S. (2014), "A novel approach for breast cancer detection using data mining techniques", International Journal of Innovative Research in Computer and Communication Engineering, 2(1), 2456-65.
- [10] Francesca Pentimalli (2017), "Updates from the TP53 universe", Cell Death and differentiation, doi:10.1038/cdd.2017.190.
- [11] Marcilio C. P., Natal, Brazil, Daniel S. A. de Araujo, Natal, Brazil Ivan G. Costa, Rodrigo G. F. Soares, Teresa B. Ludermir, Alexander Schliep (2008). "Comparative study on normalization procedures for cluster analysis of gene expression datasets", 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), DOI: 10.1109/IJCNN.2008.4634191.
- [12] D.Shanthi, Dr.G.Sahoo and Dr.N.Saravanan (2009), "Designing an Artificial Neural Network Model for the Prediction of Thrombo-embolic Stroke", International Journals of Biometric and Bioinformatics (IJBB), Volume (3) : Issue (1).
- [13] Dongare A., Kharde R. and Amit D. (2012), "Introduction to Artificial Neural Network", International Journal of Engineering and Innovative Technology (IJEIT), Vol. 2, Issue (1).
- [14] R Aruna Florence, Srikanth Bethu, V Sowmya, Kollu Anusha, and B Sankara Babu (2018). "Importance of Supervised Learning in Prediction Analysis", Periodicals of Engineering and Natural Sciences, Vol.6, No.1, 201-214.
- [15] Abbas H. Taqi (2015), "LEARNING RATE COMPUTATION FOR THE BACKPROPAGATION ALGORITHM", International Journal of Mathematics and Computer Applications Research (IJMCAR), Vol. 5, Issue 5, 65-72.
- [16] Massimo Buscema (1998), "Back Propagation Neural Network", Substance Use & Misuse, 33(2), 233-270.
- [17] Clemens A. Brust, Sven Sickert, Marcel Simon, Erik Rodner, and Joachim Denzler (2016), "Evaluation of QuickProp for Learning Deep Neural Networks", arXiv:1606.04333v2.
- [18] <https://www.alyuda.com>, under the Registration #: B276-2473-3174-Y5B6.
- [19] Md. Milon Islam, Hasib Iqbal, Md. Rezwanul Haque, and Md. Kamrul Hasan (2017), "Prediction of Breast Cancer Using Support Vector Machine and K-Nearest Neighbors", 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, Bangladesh.

BIBLIOGRAPHY OF AUTHORS



Deeman Yousif Mahmood received her BSC in Computer Science from University of Kirkuk (2008), Kirkuk-Iraq and MSC in Computer Science from University of Sulaimani (2014), Sulaimani-Iraq. She is a Lecturer from 2017 at University of Kirkuk and she is a member at IEEE and European Association for Theoretical Computer Science. Her research interest in Data Mining, Artificial Intelligence and Network Security.



Prof. Dr. Ayad Ghany Ismaeel received MSC in computer science from the National Center of Computers NCC- Institute of Postgraduate Studies, Baghdad-IRAQ at May. 1987, and Ph.D. Computer science in qualification of computer and IP network from University of Technology, Baghdad- IRAQ in 2006, visit URL: goo.gl/gn8pK4 . Currently a president of Al-Kitab University-Kirkuk-IRAQ, he was a dean of Alkitab University College-Kirkuk-IRAQ till Sep. 2018, visit URL: <https://www.uoalkitab.edu.iq/eng>, and he is professor at 7 Aug. and he is teaching and supervising undergraduate and postgraduate students in many qualifications like: Computer Science, Information Systems Engineering, Bioinformatics, Biotechnology, Cloud Computing, Healthcare/Embedded Systems, Mobile Network.



Prof. Dr. Abbas Hassan Taqi received MSC in Mathematics from University of Salahaddin, Erbil-Iraq and Ph.D. of Mathematics with accurate specialization of Computational Numerical Optimization from University of Mosul, Mosul-Iraq. Currently he is the president of University of Kirkuk and he is teaching and supervising undergraduate and postgraduate students. His research interest in Numerical Optimization, Neural Networks and Mathmetical Statistics.
