INTERNATIONAL JOURNAL OF BIOLOGY AND BIOMEDICAL ENGINEERING
DOI: 10.46300/91011.2022.16.16 Volume 16, 2022

A Robust Feature Extraction and Deep Learning Approach for Cancer Gene Prognosis

Kamala Kumari P[1], Joseph Beatrice Seventline[2]
[1]Deptartment of Electronics and Communication Engineering, Muffakham Jah College of Engineering and Technology, Hyderabad, India
[2]Department of Electronics and Communication Engineering, GITAM University, Visakhapatnam, India

Received: July 11, 2021. Revised: November 28, 2021. Accepted: December 20, 2021. Published: January 4, 2022.

Abstract— Mutated genes are one of the prominent factors in origination and spread of cancer disease. Here we have used Genomic signal processing methods to identify the patterns that differentiate cancer and non-cancerous genes. Furthermore, Deep learning algorithms were used to model a system that automatically predicts the cancer gene. Unlike the existing methods, two feature extraction modules are deployed to extract six attributes. Power Spectral Density based module was used to extract statistical parameters like Mean, Median, Standard deviation, Mean Deviation and Median Deviation. Adaptive Functional Link Network (AFLN) based filter module was used to extract Normalized Mean Square Error (NMSE). The uniqueness of this paper is identification of six input features that differentiates cancer genes. In this work artificial neural network is developed to predict cancer genes. Comparison is done on three sets of datasets with 6 attributes, 5 attributes and one attribute. We performed all the training and testing on the Tensorflow using the Keras library in Python using Google Colab. The developed approach proved its efficiency with 6 attributes attaining an accuracy of 98% for 150 epochs. The ANN model was also compared with existing work and attained a 10 fold cross validation accuracy of 96.26% with an increase of 1.2%.

Keywords— DNA sequences, Numerical mapping, Feature extraction , Artificial Neural Networks

I. Introduction

DESPITE decades of progress in medical diagnosis in terms of cures and predictions for various diseases, cancer remains one of the most fatal diseases. According to the latest reports and statistics from WHO, the death rate in India due to cancer is 79 per 100,000 deaths, which comes up to 6% of the worldwide cancer related deaths. There are about 100 types of cancer, out of which 14 are very common. In recent years, the death rate and types of invasive cancer have confirmed this issue. Mutations cause cancer. These mutations (changes) occur in genes that regulate how our cells function, particularly how they divide and expand. Cancer is a genetic disease caused by missense mutations at the intracellular level of DNA sequences. Each human consists of approximately 37 trillion numbers of cells. Each cell contains compactly bind chromosomes. These chromosomes are composed of nucleosome that is further composed of DNA wrapped around Histone protein. A gene is basically a stretch of DNA coded in the form of nucleotides Adenine(A), Thymine(T), Guanine(G),and Cytosine(C). At the molecular level DNA acts as genetic material for most of the organisms. But for virus like SARS, RNA acts as genetic material. Mutations occur in genes that regulate how our cells function, particularly how they divide and expand.

There are three factors responsible for mutations. Firstly, when a cell divides and copies its DNA to form two new cells, random mistakes in DNA occur. Every day, millions of times in everyone's body, this regular process occurs, and it usually causes no harm. However, when such errors damage a vital gene rather than a trivial gene, cancer can result. According to research, these flaws produce approximately two-thirds of cancer-causing mutations. Secondly, mutations are also caused by environmental factors like harmful rays, chemicals, nitrate, and infectious agents. Thirdly, cancer mutations can also be passed down through the generations. Fig 1 shows various causes of mutated DNA. Some people are born with mutations that increase their chances of developing cancer. Researchers claim that if these mutations occur in the coding regions of DNA, then they are most likely to turn into disease causing genes. In many cancer investigations, a powerful and successful strategy of focusing on the coding genes of greatest interest.

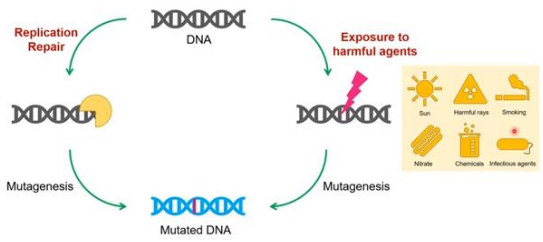

Figure 1. Mutated DNA

Conventional clinical tests are considered invasive and detrimental to the human body. There are a number of existing researches in the literature that identify cancer through image datasets. Different deep learning or machine learning algorithms were employed in these image-based methods.[1,2] These images, which are produced by X-ray or CT scanning devices, have the potential to cause lifelong damage to humans. As a result, some patients may be reluctant to acquire X-ray or CT scanning, which might be considered a drawback to image-based studies. As a result, researchers in cross-disciplinary technologies are working on non-invasive methods to predict cancer-related genes. Researchers from the genomic signal processing and machine learning fields are exploring various aspects of metadata, including dataset classification, feature extraction and prognosis of genetic diseases. In the cancer prediction dataset, feature extraction is vital task. As a result, detection algorithms should be able to identify whether a gene is malignant or not. The two main pillars of artificial intelligence are deep learning and pattern recognition. In this study, pattern recognition for genomic sequences is done through genomic signal processing algorithms. The obtained features are used to develop a deep learning model to distinguish between cancerous and non-cancerous genes.

Emerging as a predominant application of genomic signal processing, researchers started exploring for features to differentiate cancer genes. In this regard Sathapathi et al [3] is considered as one of the first empirical study to find discernment features in cancer cells by using digital IIR low pass filter with Butterworth approximation. In this study, the values of mean amplitude and mean frequency of the Power Spectral Density of the coding sequence are the discriminating parameters.

Another significant methodology was introduced by Roy and Barman [4].They discovered a particular type of ANN called FLANN for identification of cancer gene. In this study, the FLANN based adaptive filter successfully separated the cancerous genes with healthy ones with average NMSE as the differentiating parameter at sub molecular level. Related to existing methodologies the developed DSP framework showed prominent outcome identification of diseased gene over a wide range of DNA sample sequences. These researches are mostly focused on feature extraction that leads to a deep learning perspective.

In recent times, researchers have introduced deep learning to classify cancerous and noncancerous genes. Liu et al [5] used DWT based methods to identify patterns and machine learning to classify cancer genes. This study involved evaluation of statistical parameters to classify cancer and noncancerous genes with support vector machine. In contrast to other approaches, the framework distinguishes between cancer types such as breast, lung, and ovarian cancer. In this literature, SVM classifier is tested on 70 DNA samples of various cancer and non-cancerous genes and achieved better performance. Moreover, the performance can still be improved with larger datasets with other supervised Machine Learning (ML) classifiers.

Khodaei et al [6] proposed a hybrid approach to introduce machine learning into cancer prediction. In their work, Markov chain model is deployed for feature extraction and SVM as classifier. The drawback of this technique is the imbalance dataset. Khodaei et al [7] introduced a pattern recognition model to differentiate cancerous genes via machine learning algorithms. In their study, peak value versus standard deviation is taken as the discriminating feature that in turn applied to a machine-learning model. Comparison of results is done for SVM classifier with different kernels. One feature parameter discriminating cancer and non-cancerous is the limiting factor for the performance of the classifier.

Le [8] used machine learning based approach for diseased gene prediction. It presented a roadmap to predict diseased gene using 12 machine learning methods including unary, semi supervised and ensemble learning methods. Unary and semi-supervised classifiers are deployed to define the non-disease gene and ensemble methods to predicted the diseased gene. However, disease gene prediction has evolved into a broader concept that encompasses all general diseases.

Auslander N et al [9] reviewed the recent development methods that incorporate ML frameworks from molecular evolution. This study addressed the difficulties faced by the methods involving clinical applications. It strongly recommended the necessity of developing non-invasive bioinformatics frameworks. They presented a comprehensive analysis of those tools. This study presented a detailed survey on the bioinformatics area, ML methods and bioinformatics tools.

The number of features in most research is extremely limited, and as a result, some new features must be generated. The paper aims to investigate more biological discriminating features to identify deceased genes using signal-processing methods. Furthermore, it is extended to design a Deep learning model to classify cancerous genes from the dataset. In addition, Deep learning algorithms for detecting cancer genes from various sets of input features are compared in this work. The uniqueness of this paper is identification of six input features that differentiates cancer genes and an own dataset of 200 samples. These features are derived from signal

processing model in which five features are the statistical parameters of Power spectral density (PSD) of coding sequence (CDS). Last one is the Normalized mean square error (NMSE) of Adaptive Functional Link Network (AFLN).

The paper is organized as follows: Models for feature extraction employing digital signal processing techniques are described in Section 2. It also describes the deep learning algorithms used in this study. Section 3 presents the obtained results pertaining to feature extraction and deep learning algorithms. Comparative analysis is also presented in this section. The last section presents the conclusion of the work.

II. Materials and Methods

The research procedure is represented in the block diagram shown in Fig 2. Initially, the DNA sequences are downloaded in the form of a FASTA file from GenBank based on research requirements. Secondly, using DNA mapping, the alphabetical sequences are converted into a suitable format. During the feature extraction phase, digital signal processing techniques are employed to extract features. Then Classification is done through selected deep learning algorithms. Finally, the classifiers' performance is evaluated.

A. DNA Mapping

It is a prerequisite to convert the biological DNA sequence into discrete sequence using a mapping scheme before applying DSP algorithms. Voss mapping, which maps the four protein coding nucleotides A, C, G and T into four binary indicator sequences, is one of the earliest and most widely used mapping techniques[10]. EIIP is the other mapping technique that is extensively used[11]. In EIIP technique A, T, G, C is assigned with 0.1260, 0.1335, 0.0806 and 0.1340 respectively. In this work, integer-mapping techniques employed. In this mapping the four nitrogenous bases were mapped to the four integers values as follows A=2, T=0, G=3, C=1. This method incorporates DNA structural properties such as purines (A, G) > pyrimidines(C, T) that introduces bias in the DNA sequence analysis.

B. Signal Processing for Feature extraction

To extract features that distinguish cancer from non-cancerous genes, a number of Digital signal processing (DSP) methods and algorithms have evolved [12,13]. In the proposed method, five features are the statistical parameters of power spectral density (PSD) of the coding sequence (CDS). The sixth one is the Normalized mean square error (NMSE) of the Adaptive Functional Link Network (AFLN). Therefore, six features forms the attribute set for the deep learning classifiers.

1. Feature Extraction using PSD technique

The DNA string is converted into discrete sequence x[n] using integer mapping technique. Discrete Fourier Transform (DFT) of the discrete sequence is obtained using Eq(1)

$$X(k) = \sum_{n=0}^{L-1} x(n) e^{-j\frac{2\pi}{L}kn} \qquad (1)$$

$$where\ k = 0,1,2..L-1$$

The Power Spectral density of the coding sequence(CDS) is obtained using Eq(2)

$$P_s(k) = \sum_{n=0}^{N-1} |X_s(k)|^2 \qquad (2)$$

PSD plots of Coding Sequences (CDS) are investigated for identification of cancer and non-cancer genes. Spectral characteristics of CDS can be further improved by using a IIR Low pass filter with Butterworth approximation. The filter's goal is to reduce noise in the power spectrum, allowing for more accurate cancer gene prediction. It can be observed from the PSD plots that in cancer gene spikes are detected in the power spectrum plot which are lacking in normal genes. This prominent feature results in variations in the evaluated statistical parameters. This discriminative feature extraction is performed by signal processing model shown in Fig 3.

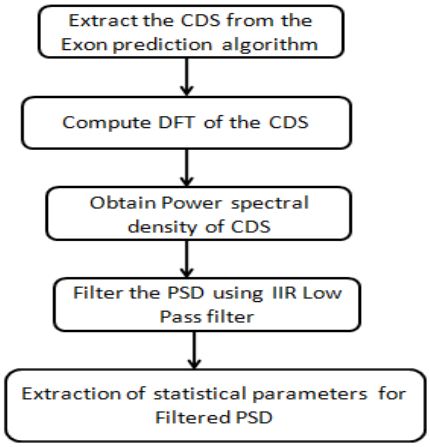

Figure 3. Feature extraction block using PSD technique

2. Feature Extraction using AFLN

AFLN model is a modified version of Adaptive exponential Functional Link Network (AEFLN) proposed by [14,15]. AEFLN is deployed for various non-linear applications such as system identification, echo cancellation, noise control.

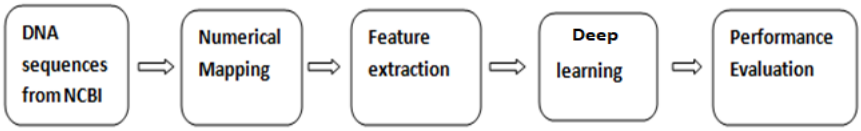

Figure 1. Block diagram of proposed work

In this study, AEFLN is modified to evaluate NMSE as the feature metric. In AEFLN, the FLN block expands the input by taking the exponent of the input signal. Unlike in AEFLN, AFLN expands the input signal by using the sinusoids alone.

The detailed process flow is shown below
Step 1: Initialization of weight vector $w(k)=0$
Step 2: Generate the tap delay input
$x(n)=[x(n),x(n-1)\ldots..x(n-N+1)]^T$, where N is the length of input.Input is expanded to M=N(2B+1)+1 terms using a non linear trigonometric functions with order B=2 as follows

$$f(k) = [1,\quad x(k),\ \sin(\pi x(k)), \cos(\pi x(k)), \sin(2\pi x(k)), \cos(2\pi x(k)), x(k-1), \sin\sin(\pi x(k-1)), \cos(\pi x(k-1)), \sin(2\pi x(k-1)), \cos(2\pi x(k-1)) \ldots\ldots x(k-N+1), \sin\sin(\pi x(k-N+1)), \cos(\pi x(k-N+1)),\ \sin(2\pi x(k-N+1)), \cos(2\pi x(k-N+1))] \quad (3)$$

Step 3: Estimation of weights for the kth input
$$w(k) = [w_0(k), w_1(k), w_2(k) \ldots\ldots\ldots\ldots w_{M-1}(k)]^T \quad (4)$$
Step 4: estimate output using the following relation
$$y(k) = f(k)^T . w(k) \quad (5)$$
Where y(k) is the output of FLN block.
Step 5. The desired output is evaluated from the below mentioned input output relation
$$d(k) = 2\left\{\frac{1}{1+e^{[-uv(k)]}} - \frac{1}{2}\right\} \quad (6)$$
$$\text{where}, v(k) = 1.5\,x(k) - 0.3\,x^2(k) \quad \text{and } u = \begin{cases} 4 & v(k) > 0 \\ 0.5\,v(k) & \leq 0 \end{cases}$$
Step 6 : Estimated error is evaluated as
$$e_s(k) = d(k) - y(k) \quad (7)$$
Step 7 : weight matrices for LMS algorithm are updated using $w(k+1) = w(k) + 2\alpha\, e_s(k) f(k)$ (8)
where α is the learning rate
The schematic for the above process flow is depicted in Fig.4 Here input x(k) is mapped DNA sequence which is discrete in nature. Using the Non-linear trigonometric expansion, the given input is mapped to a higher dimension space using the Eq(3) gives a clear idea about the gene characteristics.. Initially the weights are set to zero. For each iteration , the next weight vector is estimated, and the estimation error is calculated by adding the term $2\alpha\, e_s(k) f(k)$ to the current weight estimate as shown in Eq(8). Output is estimated as the product of higher dimensional input vector and updated weight vector. The desired output is generated using the Eq(6). Error value is evaluated as difference between actual input and desired input and weights are adjusted to reduce the error value. The normalized mean square error(NMSE) value for the last 100 iterations of the above algorithm is evaluated to obtain the feature value for classification of cancerous cell.

C. Deep Learning for Cancer Gene Prediction

Artificial Neural Networks (ANN) is a deep learning algorithm that is emerged as software model of biological neurons that are based on natural biological systems[16].

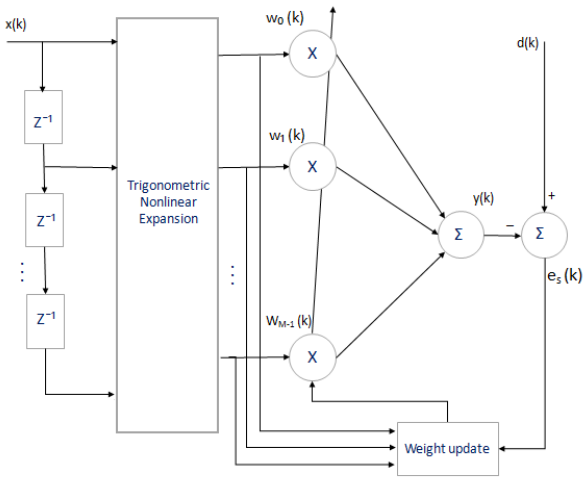

Figure 4. Feature extraction block using AFLN

Artificial neurons follow the workflow of biological neurons. A typical biological neuron consists of dendrites, axon, soma and synapses. These resemble a mathematical model that produces sharp electrical spikes. These spikes are transmitted along the axon and synapses from the sending neuron to many other neurons. Similarly, a set of interconnected neurons is referred, as an artificial neural network with each connection link is associated with a weight that has information about the input signal. These interconnected neurons are grouped in three layers in the basic architecture: input, hidden and output layers[17]. Each neuron has an internal state called an activation signal. The input signals are combined with the activation rule to create output signals that can be directed to other units. The three building blocks of ANN are network topology, adjustment of weights and activation functions. The two types of ANNs are perceptions, which are the simplest forms of ANNs used for binary classification, and multilayer ANNs, which are a more advanced version of perception used to tackle complex classification and regression issues. For identification of cancer genes, we have used multilayer ANNs.
The mathematical representation of each neuron is given by
$$f(x) = \sum_{i=1}^{n} X_i W_i + bias \quad (9)$$
Where X_i is the input value and W_i is the weight from input to output layer.
To match to the exact output values, activation functions are applied. Following are some activation functions commonly used in artificial networks
Step function f(x) = 1 $x \geq 0$ (10)
Sigmoid function $f(x) = \frac{1}{1+e^{-x}}$ (11)
Rectified Linear Unit function $f(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases}$

In the ANN's architecture, the number of neurons in the input layer is equal to the number of features in the dataset. The network also has a hidden layer, with the number of hidden layers being counted as one layer. The cost function

calculates the error value, which is the difference between actual and anticipated values, after forward propagation.

D.Experimental Setup

We applied deep learning approach ANN to predict whether a cell is cancer or non-cancerous using Python programming language. We used HP pavilion with 4.7 GHz (Turbo Frequency) Quad-core Intel Core i7-1165G7 11th Gen Processor and 8GB RAM. On the python notebook, all of the necessary libraries were installed and used for data analysis. Feature extraction is implemented in MATLAB 2016b.

E. Dataset

The National Centre for Biotechnology Information (NCBI) also called, as Genbank is one of most prominent databases for various genetic diseases. A set of 200 genes associated with 140 cancerous and 60 non-cancerous genes are extracted from NCBI Genbank https://www.ncbi.nlm.nih.gov/.

The sequences in the dataset are also mentioned in the previous studies. The features that differentiate cancer genes are Mean, Median, Standard Deviation, Mean Deviation, Median Deviation and NMSE are shown in Table 1 and Table 2. The above stated features are evaluated for all 200 DNA sequences to form the dataset. The cancerous genes are identified as negative class and non-cancerous as positive class. The back propagation method was used to train and test the ANN models on the TensorFlow backend using the Keras package in Python for the collected dataset.

F. Performance Metrics

The effectiveness of machine learning algorithms is assessed using a set of performance metrics. To evaluate the parameters, a confusion matrix including TP, FP, TN, and FN for actual data and predict data is constructed as shown in Table 3.

Table 3. Confusion matrix

		Actual	
		Cancer	Non Cancer
Prediction	Cancer	True Positives (TP)	False Positives (FP)
	Non Cancer	False Negatives (FN)	True Negatives (TN)

Table 1. Feature parameters for Cancerous cells

S.No	Accession No.	Mean	Median	Std	Mean Deviation	Median Deviation	NMSE
1	AB383726.1	1.0426	0.9765	0.5746	0.4541	0.3934	0.070512
2	AB385129.1	1.3814	1.2846	0.7383	0.5771	0.4759	0.06271
3	AF002672.1	1.0606	1.0159	0.5646	0.4385	0.3693	0.040549
4	AF005068.1	1.3957	1.3051	0.7591	0.5917	0.4942	0.062347
5	AF041259.1	1.3783	1.3015	0.7247	0.5795	0.4904	0.052702

Table 2. Feature parameters for Non-Cancerous cells

S.No	Accession No.	Mean	Median	Std	Mean Deviation	Median Deviation	NMSE
1	AF083883	0.6637	0.6195	0.3552	0.2817	0.2485	0.076
2	AF186607.1	0.7208	0.6221	0.3858	0.2977	0.2224	0.09
3	AF186613.1	0.7236	0.6486	0.381	0.2901	0.2432	0.089
4	AF007546	0.8226	0.7831	0.4263	0.3334	0.2785	0.0761
5	AF003934	0.6514	0.6252	0.3507	0.2834	0.2571	0.106

The performance metrics are evaluated using the following equations

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (13)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (14)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (15)$$

$$\text{F1 Score} = \frac{2TP}{2TP+FP+FN} \quad (16)$$

III. Results and Discussions

There are two phases to our proposed technique. The feature extraction phase comes first, followed by the evaluation of an artificial neural network model for cancer prediction. In the feature extraction phase, we have evaluated 6 attributes using two different DSP frameworks .These attributes are evaluated for a set of 200 DNA samples. The features added to our dataset are mean, median, Standard Deviation, Mean absolute deviation, median absolute deviation and NMSE. Features of cancer or non-cancerous gene are extracted in the initial phase of this study. The features are derived using PSD plots and AFLN. Fig 5 and 6 are PSD plots of cancer and non-cancer genes. It can be observed that spikes are generated in case of cancer genes alone. The PSD plot of the non-cancerous gene is very noisy.

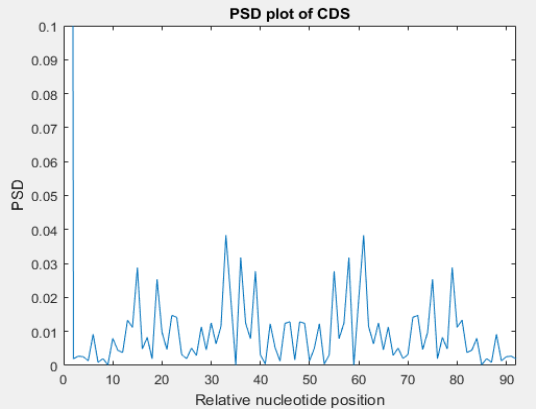

Figure 5. PSD plot for non-Cancer cell Accession No. AF186607

From the PSD plots, it is evident that there can be some parameters that can make the difference. The statistical parameters are evaluated for the same and are tabulated in Table 1 and 2. From Table 1 and 2 it clearly shows that the range of values for cancer and non-cancerous DNA sequences are quite different. AFLN model is identifies cancer genes based on NMSE values. It is required to set 1000 as the number of iterations to minimize the error and NMSE be stable. The value α is chose to be 0.002. Table 1 and 2 depicts the experimental NMSE values for cancerous and non-cancerous genes.

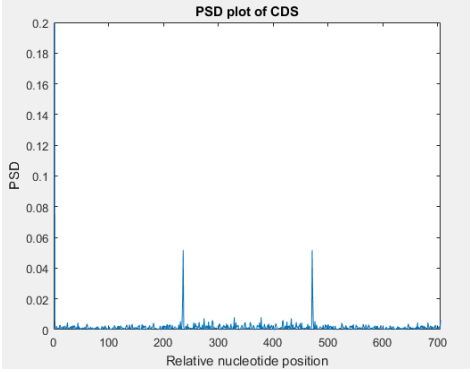

Figure 6. PSD plot for Cancer cell of accession no. AF008216

It can be observed that NMSE is less than 0.07 for cancerous and for non-cancerous this value is above 0.07.Therefore 0.07 is chosen to be the threshold to differentiate the two classes.

In our study, all the features give a significant difference between cancer and non-cancerous genes. A dataset is formed with these six features as attributes to deploy deep learning algorithms to classify DNA sequence into cancer or non-cancerous. The output is a labelled data with cancerous gene is assigned with 1 and non-cancerous as 0.

In our present work, as there are 6 attributes the input layer consists of 6 neurons. These are connected to 16 neurons in the first hidden layer. There exists a 16 to 32 mapping between first and second hidden layer. Cancer gene identification being a classification problem there exist only one neuron in the output layer. The model is tuned for 150 epochs with RELU and Sigmoid as activation functions in first and second hidden layers respectively. The architecture of ANN is illustrated in the Fig 7. The dataset is split into 75% train data and 25% test data. The ANN model is tested on three types of datasets. Firstly, the proposed dataset with 6 attributes. Secondly, the dataset with 5 features as proposed in Liu et al [5]. Thirdly, the dataset with one feature as proposed in Khodaei et al [7]. Table 4 shows the obtained accuracy for all three types of data sets for various epochs. It is observed the ANN model with 6 features attained the highest accuracy of 98% for prediction of cancer gene. Figs 8, 9 and 10 show the performance of ANN model on three types of datasets. We have also performed 10-fold cross validation for the 6-feature dataset. Table 5 shows the performance metrics for 10-fold cross validation. Precision, Recall, F1 score of the proposed ANN model are evaluated as 94%, 96% and 95% respectively. We have also compared the proposed model with existing model as shown in Table 6. Existing ANN model Khodaei et al[7] was a dataset with 750 DNA sequences with one feature. The dataset is an imbalanced dataset with 82% cancerous genes and 18% noncancerous genes. Existing model Khodaei et al[6] is a dataset of one input feature. Compared with the existing models, the accuracy of our cancer gene prediction model is increased by 1.2%.

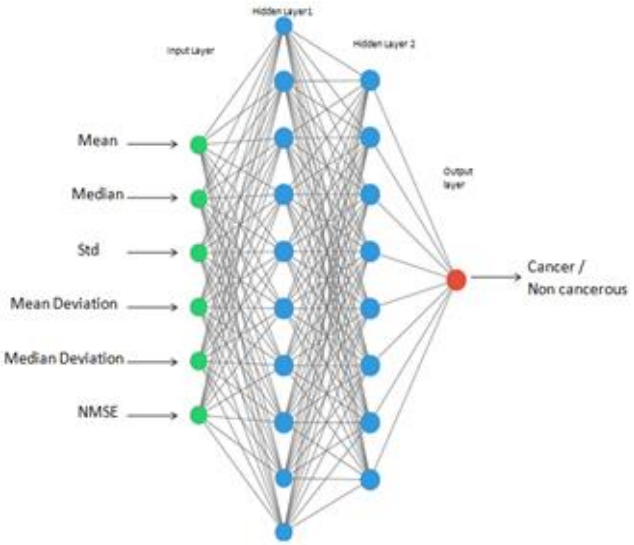

Figure 7. ANN for Cancer gene prediction

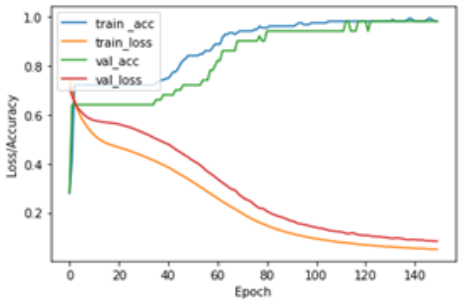

Figure 8. Training and validation accuracy/loss for 6 feature dataset

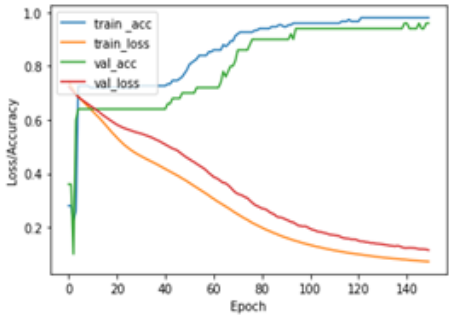

Figure 9. Training and validation accuracy/loss for 1 feature dataset.

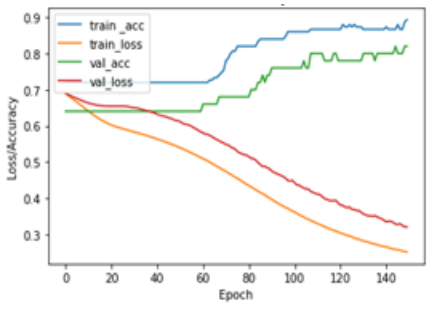

Figure 10. Training and validation accuracy/loss for 5 feature dataset

Table 4. Accuracy for various epochs

Epochs	Our dataset	Dataset with 5 features	Dataset with 1 feature
10	63%	64%	62%
30	65%	66%	64%
50	78%	72%	64%
70	80%	90%	66%
80	94%	80%	64%
100	96%	94%	80%
150	98%	96%	78%

Table 5. Performance metrics of proposed model

Performance metric	Value
Precision	94%
Recall	96%
F1 score	95%
Accuracy	96.25%

Table 6. Comparison with existing work

Method	10-fold Accuracy
Proposed	96.25%
Khodaei et al[7]	95%
Khodaei et al[6]	84%

IV. Conclusion

This paper presented a novel approach to predict cancer gene by developing a deep learning model with 6 unique attributes. We presented robust feature extraction techniques in the first phase and in the second phase, an artificial neural network model is introduced to predict cancer gene. Mutation cause cancer and if these mutations occur in exons regions leads to disease, causing mutated DNA. Power spectral density is evaluated for the coding regions. From the PSD plots, it is evident that there can be some parameters that can make the difference. Statistical parameters such as mean, median, standard deviation, mean absolute deviation and median absolute deviation are evaluated. AFLN block is deployed for evaluation of NMSE. Finally, 6 features are evaluated for 200 DNA sequences to form the dataset. The proposed cancer prediction model is an Artificial Neural Network consists of six input layer, one hidden layer, and one output layer. Finally, dataset containing 6 features is used for the evaluation of the propose ANN model. The model is tested on 3 datasets with 6 features, 5 features and one feature. In our experiment ANN model attained highest accuracy of 98% with 6 attributes. Our proposed ANN model with 6 input features boosted the accuracy of the cancer gene prediction model by 1.2 %when compared to previous work. In the future, more discriminating features can be extracted to predict cancer gene and develop machine

learning models or deep learning models that can further increase the accuracy.

References

[1] Alanazi SW, M. M. Kamruzzaman, Md Nazirul Islam Sarker, MadallahAlruwaili, Yousef Alhwaiti, Nasser Alshammari, and Muhammad Hameed Siddiqi, "Boosting Breast Cancer Detection Using Convolutional Neural Network" Journal of Healthcare Engineering ,2021

[2] Wadhwa G, and A. Kaur, "A Deep CNN Technique for Detection of Breast Cancer Using Histopathology Images," 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), 2020, pp. 179-185, doi: 10.1109/ACCTHPA49271.2020.9213192.

[3] Satapathi GN, P. Srihari, A. Jyothi and S. Lavanya, "Prediction of cancer cell using DSP techniques," 2013 International Conference on Communication and Signal Processing, 2013, pp. 149-153, doi: 10.1109/iccsp.2013.6577034.

[4] Roy S, S. Barman, A non-invasive cancer gene detection technique using FLANN based adaptive filter. Microsyst Technol 27, 463–478 (2018).

[5] Liu D, W, Run-Ping Jia, Cai-Feng Wang, N. Arunkumar, K. Narasimhan, M. Udayakumar, V. Elamaran, "Automated detection of cancerous genomic sequences using genomic signal processing and machine learning", Future Generation Computer Systems (2019)

[6] Khodaei A, Feizi-Derakhshi MR, Mozaffari-Tazehkand B. "A Markov chain-based feature extraction method for classification and identification of cancerous DNA sequences". Bioimpacts. 2021;11(2):87-99. doi:10.34172/bi.2021.16

[7] Khodaei A, Feizi-Derakhshi, Mozaffari-Tazehkand B, "A pattern recognition model to distinguish cancerous DNA sequences via signal processing methods", Soft Comput 24, 16315–16334 (2020). https://doi.org/10.1007/s00500-020-04942-4

[8] Le DH, "Machine learning-based approaches for disease gene prediction", Briefings in Functional Genomics, Volume 19, Issue 5-6, September-November 2020,

[9] Auslander N, Gussow A. B, Koonin E.V, "Incorporating Machine Learning into established Bioinformatics Frameworks", Int. J. Mol. Sci. 2021,22,2903.

[10] Massicotte D, M. A. Jaber, M. Massicotte and P. Massicotte, "Low Complexity Frequency Monitoring Filter for Fast Exon Prediction Sequence Analysis," 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2019, pp. 1-5,

[11] Dessouky AM, Abd El-Samie FE, Fathi H, Salama GM. "Efficient implementation of parametric spectral estimation techniques for DNA exon prediction". Nucleosides Nucleotides Nucleic Acids. 2020; 39(8): 1200-1221.

[12] Gayathri (2017), Analysis of genomic sequences for prediction of cancerous cells using wavelet technique. Int Res J Eng Technol 4(4):1071–1077

[13] Salmi N and Zuherman Rustam, "Naive bayes classifer models for predicting the colon cancer" 2019 IOP Conf. Ser.: Mater. Sci. Eng. 546 052068

[14] Patel V, V. Gandhi, S. Heda and N. V. George, "Design of Adaptive Exponential Functional Link Network-Based Nonlinear Filters," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 63, no. 9, pp. 1434-1442, Sept. 2016, doi: 10.1109/TCSI.2016.2572091.

[15] Patel V, S. Subhra Bhattacharjee and N. V. George, "Convergence Analysis of Adaptive Exponential Functional Link Network," in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 2, pp. 882-891, Feb. 2021, doi: 10.1109/TNNLS.2020.2979688.

[16] Joshi C, Ranjeet Kumar Ranjan, Vishal Bharti, "A Fuzzy Logic based feature engineering approach for Botnet detection using ANN", Journal of King Saud University - Computer and Information Sciences,2021,

[17] Kazi M-K, Eljack, F., Mahdi, E., "Predictive ANN Models for Varying Filler Contentfor Cotton Fiber/PVC Composites based on Experimental Load Displacement Curves", Composite Structure ,Composite structures ,2020

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en_US