ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 7, July 2017

TLBO algorithm for the Design of Low Pass IIR Digital Filter for Prediction of Cancer Cell

Baldeep Singh[1], Damanpreet Singh[2], Amandeep Kaur Brar[3]

PG Student [ECE], Dept. of ECE, Punjabi University, Patiala, Punjab, India[2]

Associate Professor, Dept. of CSE, Sant Longowal Institute of Engineering & Technology, longowal, Punjab, India[1]

Assistant Professor, Dept. of ECE, Punjabi University, Patiala, Punjab, India[3]

ABSTRACT: This paper develops an innovative methodology for prediction of cancer cell from homo sapien's DNA. Even before pre-mature stage of cancer, mutation in DNA can be detected. DSP approach used for prediction of cancer cell identifies coding and non-coding regions. This approach uses DFT, PSD and then use of optimization algorithm named Teaching Learning Based Optimization(TLBO) with 3rd order of filter. Ratio of mean amplitude and frequency differentiates cancer from non-cancer cell. With low order of filter using TLBO, higher accuracy and efficiency can be achieved as compared to butter worth filters. Ratios having value less than 1 predicts cancer cell. This research is helpful in cancer's treatment and for making anti-cancer drugs.

KEYWORDS: Exons, Discrete fourier Transform(DFT), Power spectral Density(PSD), Teaching learning based Optimization (TLBO), IIR-LPF (infinite impulse response Low pass filter)

I. INTRODUCTION

Genomic signal processing (GSP) deals with processing genomic signals that includes signals originated from DNA sequence and mRNA sequence. Cell is a fundamental unit of living being and every instruction is contained in the chemical base of DNA chain. Goal of GSP is to use these signals to classify cancer on molecular basis and to screen these genes. Cancer is one of the dreaded diseases causing death all over the world. Abnormality in genetics of cell is the major cause of cancer. These abnormalities are inherited or sometimes it occurs through errors in DNA replication. Nucleotide sequence change or mutation is the major reason behind cancer. These mutations occur through cell division or in response to some injuries. Exposure to environmental agents such as radiation or chemicals causes mutations. Spectral characteristics of coding regions are analyzed for cancer and non -cancer cells.[6]

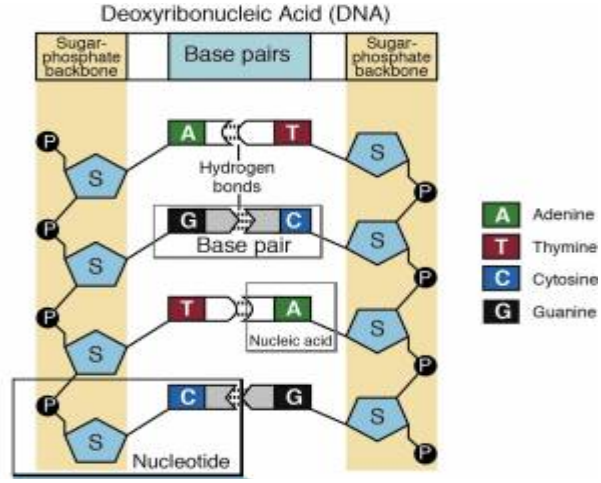

Fig.1 Structure of DNA having hydrogen bonds between nucleotide bases

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 7, July 2017

DNA is a double helix structure which contains two complementary strands of sugar phosphate with nucleotide bases attached to it. Nitrogen base, pentode sugar, and phosphate group are the three major components of nucleotides. Adenine (A), thymine (T), Guanine (G), and Cytosine (C), these four nucleotides executes cell's activity. T is always paired with A and C is always paired with G. A and T are bonded by double whereas G and C are bonded by triple bond. Therefore, A-T forms weak and G-C forms strong bond. Due to some environmental agents such as radiations or some chemicals, sometimes insertions, deletions or substitutions in these bonds occur. Thereby causing replacement of these bonds which causes cancer.[6]

Frequency analysis of character string concluded that Numerical values are assigned to amino acids by modeling protein coding as an FIR digital filter in which Character strings are mapped into numerical sequences. Digital filtering approach was used for translating proteins from nucleic acids. The computational techniques with visual tools synergistically complement "character-string-domain"[1]. Different biomedical devices are used in Nanotechnology for cancer prediction. To identify novel genes, nanoscale biomarkers are used which defines functions and regulatory networks. Toxicity of Nano materials was reduced by coating them with biocompatible materials. Size of biomarker is important in penetrating the porous structure of cell membrane. Challenge for nanotechnology is designing cancer treatments and then their advancement from methods of treatment to prevention. Interdisciplinary training and non-trivial translation of these biomarkers from cellular GSP to human and animal applications are another challenges. [2]

Major challenge in predicting the cancer is maximization of prediction accuracy of gene locations and coding regions. Exon prediction methods and novel signal processing based gene techniques are used for exon and gene prediction problem. Symbolic to numeric representations of DNA are presented and then a technique for acceptor splice's site is represented which combines these two methods. This hybrid technique consistently reduces the false positives at sensitivity's different levels [3]. Electron ion interaction potential (EIIP) values is another technique used forAnalysis of Genomics and Proteomics. Physico-mathematical approach named resonant recognition model (RRM) having two stages interprets protein sequence linear information. First stage converts amino acid to numerical sequence and second one analyses original numerical sequence using FFT. [4]

For Finding Codon Bias in DNA Sequences, Period-3 component identification using window and Filtered Spectral Rotation method (From gaussian distribution) are used. DSP based model explains period-3 component's mechanism and relates its identification to detection of nucleotide bias in coding region. This model completely characterizes DNA spectrum by filtered polyphase sequences. Standard signal processing tools such as digital filtering of the sequence enhances the codon bias's detection.[5]

Different DSP algorithms are used for analysis and DFT, DWT, entropy and parametric modelling are the key factors. IIR and FIR filtering techniques having magnitude of basic ideal filter with frequency responses be high pass, low pass, band pass or band stop are used. For gene prediction in cancer cell, IIR anti notch filters are used which provides high accuracy with a slight number of increase in multipliers. FIR filters are used in limited window length between 351 and 651. Long window length gives poor domain resolution implying more computations whereas short window length increases the level of noise. Welch's periodograms method is used for stationary random signals to obtain power spectral density in frequency domain whereas for non- stationary signals, Short time fourier transform (STFT) is used. Autoregressive, Moving average and combination of both are three different parametric models used for random sequence analyzation that depends on characteristics of PSD.[6] Principal Component Analysis (PCA) Technique take eigen values and vectors in which Correlation Coefficients are calculated. Positive correlation gives normal and negative gives cancer cell. This technique is based on maximum variance between samples rather than maximum correlation between samples, thereby providing higher accuracy. [7]

With the help of the digital filters, removal of noise and bandwidth limitation is possible. Infinite impulse response filters are generally used because of their less number of coefficient requirement for same frequency response. Due to less efficiency in filter structure and coefficient quantization errors, transformations techniques are not preferred. Optimization techniques having different evolutionary algorithms are employed for filter design. Teaching learning based optimization is such an algorithm which is based on teaching and learning process used to design 3rd order LPF.[11,12].The paper is organized as follows: SectionIIdescribes the basic DSP technique used in prediction of cancer cell. Design of LPF is discussed in sectionIII. Detailed Algorithm and its analysis is discussed in section IV. In section V,results are discussed and compared. Last section concludes the whole research.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 7, July 2017

II. DSP TECHNIQUE

There are some signals that are not revealed in time domain. Such signals belong to frequency domain which makes use of discrete fourier transform [9]. Genomic sequences in DNA are converted into numerical sequences with the use of several mapping techniques [10]. One of the new mapping technique is based on the strong and weak hydrogen bonds between nucleotide bases. On the basis of strong and weak bond, one binary sequence is generated for each nucleotide. A,T, C & G are four nucleotide bases .DNA string of length N is defined by x[n] with alphabets. Let us define a single binary indicator sequence $x_s[n]$.

If x[n] = T A A C G A G T C, then $x_s[n]$ = 0 0 0 1 1 0 1 0 1

The DFT Xs[k] of binary sequence is

$Xs[k] = \sum xs[n] e^{-j2\pi nk/N}$ (1)

k=0,1,2,……..N-1 and n=0,1,2,…….N-1.

Then the Power Spectral density of the sequence is given by

$Ps[k] = \sum |Xs[k]|^2$ (2)

Plot of Ps[k] of coding region is the indicator of cancer cell or non- cancer cell. These spectral characteristics of DNA can further be improved by the use of optimization algorithms with different order of filters. These optimization algorithms provide better prediction of cancer cells.

III. DESIGN OF LPF USING TLBO

In this section the application of TLBO for the design of IIR LPF is explained. TLBO has two phases named teacher phase and learner phase. Results achieved by learners in class not only depend on the knowledge given by the teacher but it also depends upon their own quality. Teacher shares his knowledge with learners in order to improve their academic performance. Learners also gain knowledge by discussing different matters among others.[11] Population in IIR LPF using TLBO is analogous to L number of learners in class. Fitness value is analogous to student scores. Difference equation for IIR digital filter is:

$$y(n) = \sum_{k=0}^{\infty} h(k)x(n-k) \tag{3}$$

$$y(n) = \sum_{k=0}^{N} b_k x(n-k) - \sum_{k=1}^{M} a_k y(n-k) \tag{4}$$

where h(k) represents filter's impulse response, x(n) gives discrete input whereas y(n) gives discrete output, and a_k ,b_k defines coefficients of filter. IIR filter's transfer function is given by-

$$H(z) = \frac{\sum_{k=0}^{M} a_k z^{-k}}{1 + \sum_{k=1}^{N} b_k z^{-k}} \tag{5}$$

Cascading of first order and second order is used for the design of third order filter with following equation-

$$H(w,x) = A\left(\prod_{u=1}^{M} \frac{1 + p_{1u}e^{-jw}}{1 + q_{1u}e^{-jw}}\right)\left(\prod_{v=1}^{N} \frac{1 + g_{1v}e^{-jw} + g_{2v}e^{-2jw}}{1 + h_{1v}e^{-jw} + h_{2v}e^{-2jw}}\right) \tag{6}$$

Where $x=[p_{11},q_{11}….g_{11}…h_{2N},A]^T$ and vector x gives the coefficients of filter having dimension S×1 with S=2M+4N+1 and A defines the gain.

Design process includes minimizing the absolute magnitude response error. L_1 norm approximation of magnitude response is minimized. TLBO require few control parameters and it needs minimum tuning. Parameters remain fixed during the optimization process. [12]. L_1 norm error of magnitude response is denoted by E1(X). According to design criteria defined, the objective to be minimized is given by:

$$Minimize\ F(X) = E_1(X) \tag{7}$$

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 7, July 2017

Fitness function of i[th] user for the design of IIR LPF is derived from objective function and is given as-

$$f(X_i) = Minimize(F_i(X)) \qquad (i = 1,2 \dots, NL) \tag{8}$$

Fittest learner's function value becomes global best and corresponding marks scored by him in different subjects are set as best marks globally.[12]

IV. ALGORITHM

Find power spectral density for coding sequence of DNA.
Step-1- Input the DNA coding sample having length up to different base pairs and then map each alphabet base into one binary sequence on the basis of strong and weak bond.
Step-2- using Eq.1, DFT of DNA sequence is calculated.
Step-3- Obtain the PSD of sequence using Eq.2.
Step-4- Use TLBO algorithm for designing 3rd order LPF having following coefficients-
a=1.077239; b=-.663156;p=-.285299;q= 1.012133;r=-1.380685;s=.733563;h=.033490

Step-5-Filter obtained PSD using the optimized LPF.
Step-6-Get mean amplitude and mean frequency of filtered PSD.
Step-7- Compare PSD of cancer and non- cancer cell.

V. RESULTS AND INFERENCES

DFT power spectrum with optimization algorithm gives effective and accurate results. Different DNA databases have been taken for sample test. Ratio of mean amplitude and mean frequency clearly distinct cancer cells from non-cancer cells. It has been observed that the samples having ratio less than 1 belongs to cancer whereas others belong to non-cancer cells. Less spikes and higher accuracy has been observed. Even with less order of filter, higher accuracy is achieved compared with high order filters. TLBO algorithm provides higher performance in designing of filters and convergence rate is also high. It is observed that cancer cells are spiky whereas non-cancer cells are less spiky. Figure 3,4,5,6,7 & 8 clearly depicts DFT power spectrum.Different homo sapien samples over which this algorithm is tested are listed below-

Table1. Cancer Cells for 3rd order filter

Serial number	Accession number	Mean Amplitude	Mean frequency	Ratio
1	AF008216.1	0.2230	0.5000	0.4460
2	NM_005732.3	0.3044	0.5000	0.6087
3	NM_012403.1	0.2539	0.5000	0.5079

Table2. Non-cancer Cells for 3rd order filter

Serial number	Accession number	Mean Amplitude	Mean frequency	Ratio
1	AF083883	0.2732	0.2500	1.0927
2	AF186613.1	0.3012	0.2500	1.2047
3	AF007546	0.2865	0.2500	1.1459

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 7, July 2017

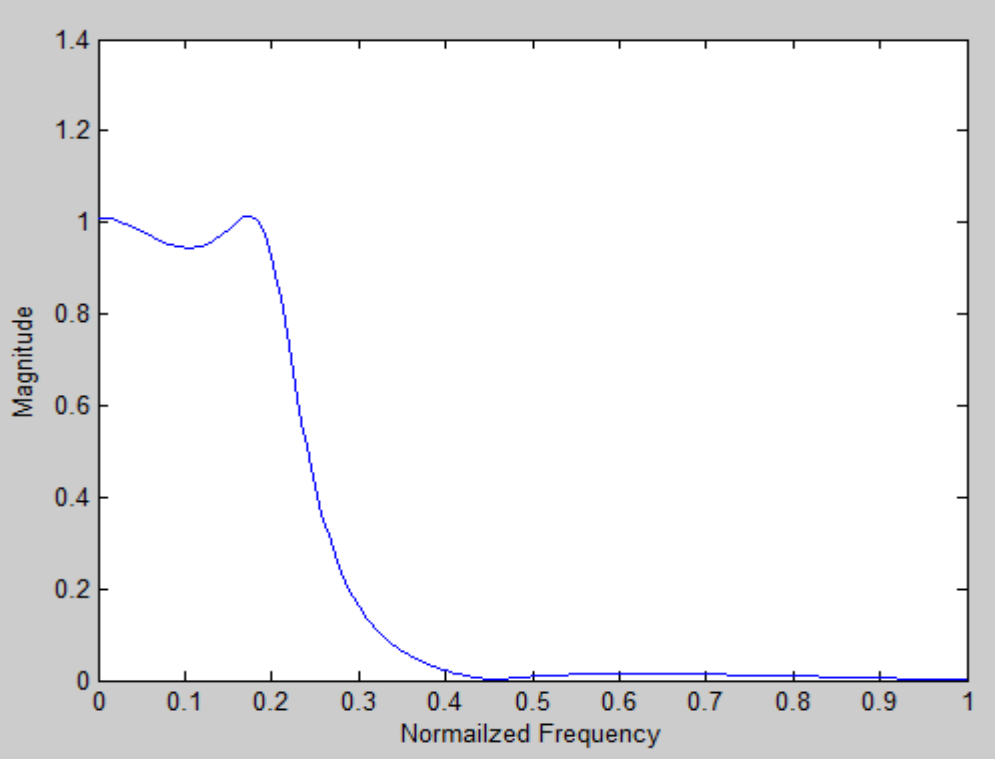

Fig.2 IIR LPF response using TLBO algorithm

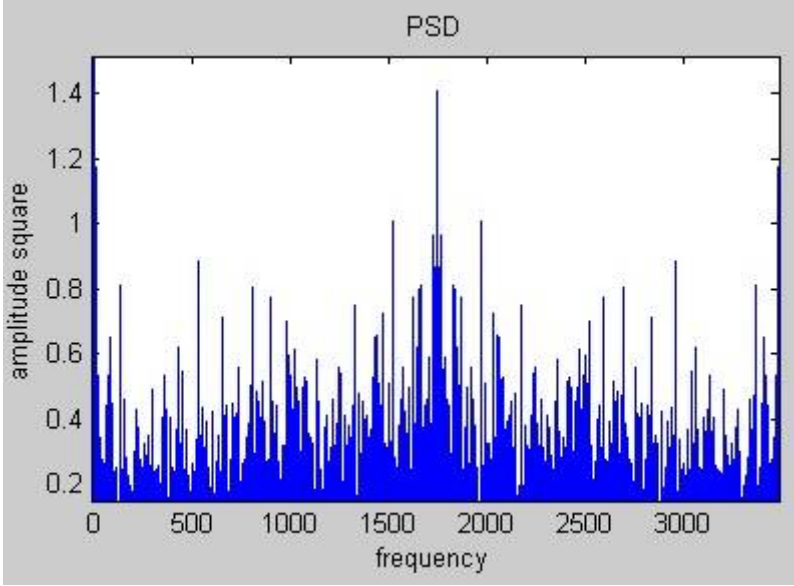

Fig.3 DFT power spectrum length for cancer cell of accession no. AF008216.1 and length of Exon 1-5785 bp with the use of optimization algorithm

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 7, July 2017

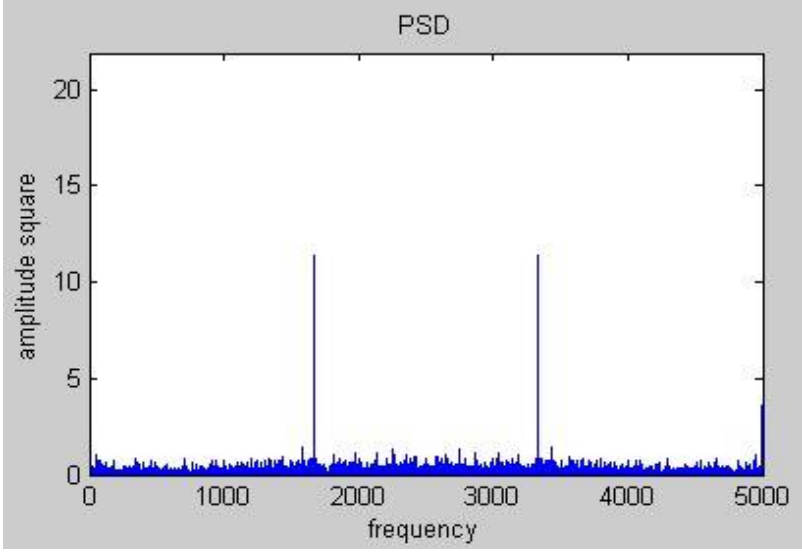

Fig.4 DFT power spectrum length for cancer cell of accession no. NM_005732.3 andlength of Exon 1-6597 bp with the use of optimization algorithm

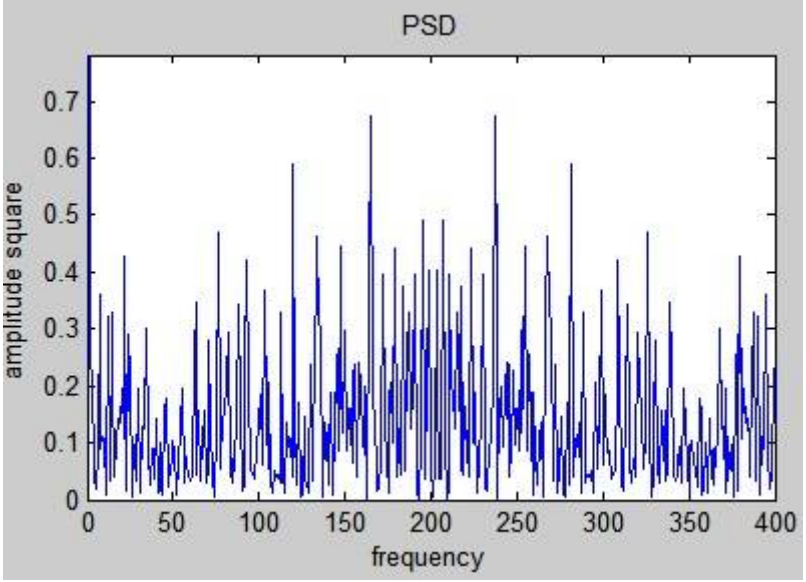

Fig.5 DFT power spectrum length for cancer cell of accession no. NM_012403.1 and length of Exon 1-705 bp with the use of optimization algorithm

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 7, July 2017

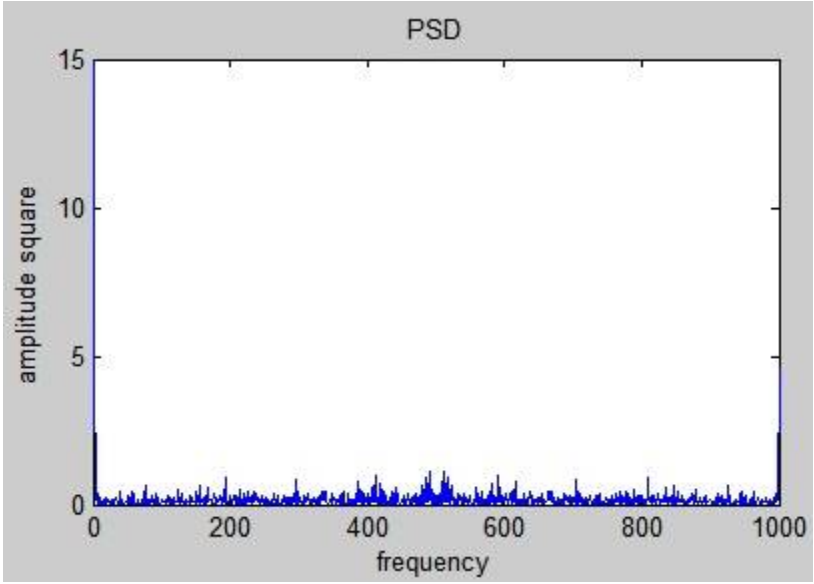

Fig.6 DFT power spectrum length for normal cell of accession no. AF083883 and length of Exon 1-1466 bp with the use of optimization algorithm

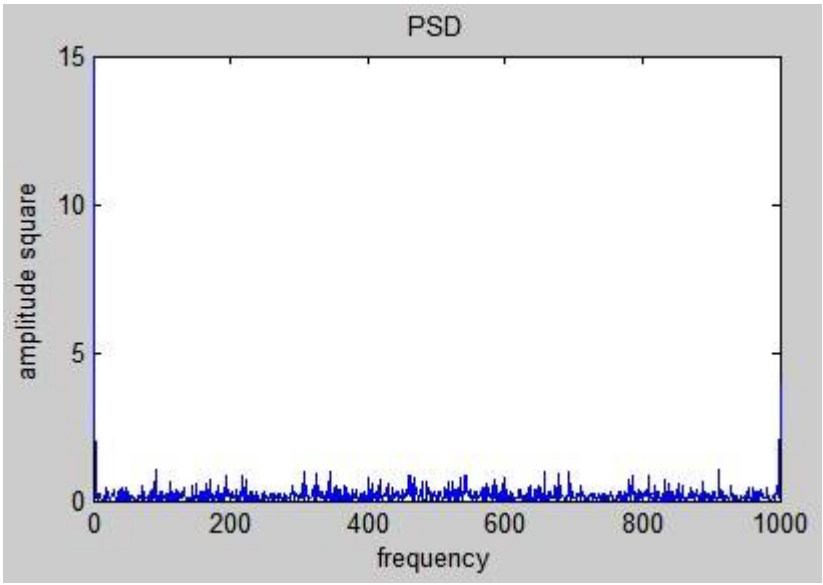

Fig.7 DFT power spectrum length for normal cell of accession no. AF186607.1and length of Exon 1-1301bp with the use of optimization algorithm

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 7, July 2017

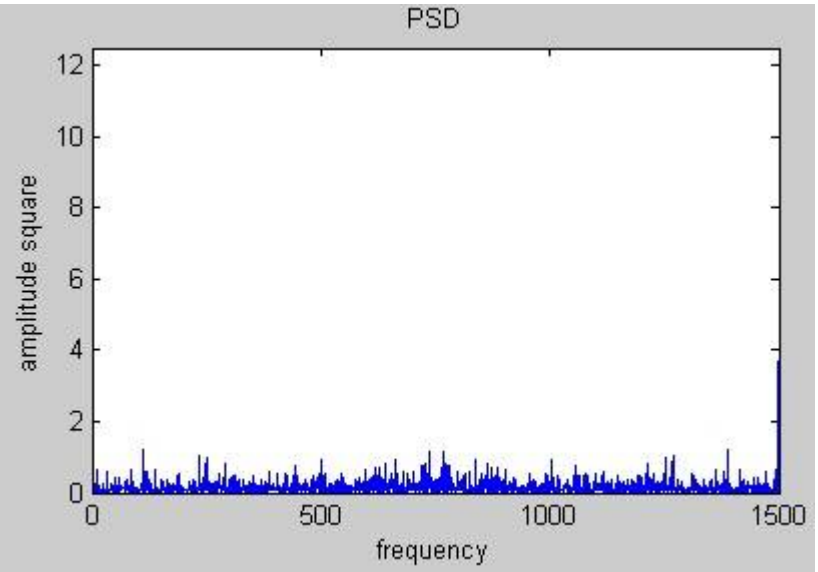

Fig.8 DFT power spectrum length for normal cell of accession no. AF007546 and length of Exon 1-2128 bp with the use of optimization algorithm

Table 3.Comparison between butterworth and TLBO algorithm LPF's

Homosapien cell	Accession number	Ratio of mean amplitude and frequency using IIR butter worth 10^{th} order filter	Ratio of mean amplitude and frequency using TLBO algorithm with 3^{rd}order filter
Cancer Cells	AF008216.1	0.52	0.4717
	NM_005732.3	0.8	0.6087
	NM_012403.1	0.52	0.5078
Normal Cells	AF083883	1.43	1.0929
	AF186613.1	3.2	1.2047
	AF007546	1.43	1.1459

VI. CONCLUSION

Here this DSP technique is tested over different samples of gene bank. Results show that this algorithm is an easy tool for prediction of cancer which helps further in treatment of cancer. Lower order filter can be designed with very low cost and power consumption for such design will be low. Even before first stage of cancer, just from DNA coding, cancer can be predicted and treated. Other efforts can be made by using FIR filters to improve the accuracy of these

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijareeie.com

Vol. 6, Issue 7, July 2017

results but window length is of great concern in such filters. Next step in this research would be to generalize it for other genes and use other filters for more accuracy.

REFERENCES

[1] Anastassiou, D: "DSP in genomics: Processing and frequency domain analysis of character strings". *IEEE, 0*-7803- 7041- 2001
[2] ChenJie and Wong Stephen T.C.: "Nanotechnology for Genomic signal Processing in cancer Research". *IEEE Signal Processing Magazine*, January, 2007
[3] Akhtar, Ambikairajah M., E. & Epps, J: "Signal Processing in Sequence Analysis: Advances in Eukaryotic Gene Prediction". *IEEE Journal of selected topics in signalprocessing*, Vol.2, No.3, June 2008
[4] Deergha.Rao K and Swamy M.N.S: Analysis of Genomics and Proteomics using DSP techniques. IEEE Transactions on circuits and Systems, Vol 55, No. 1. Pp 370-378, February 2008
[5] Tuqan. J. and Rushdi. A:A DSP based approach for finding the codon bias in DNA sequences. IEEE journal on signal processing, vol.2.No. 3, June, 2008.
[6] Inbamalar T M and Sivakumar R: Study of DNA Sequence Analysis Using DSP Techniques Journal of Automation and Control Engineering Vol. 1, No. 4, December 2013
[7] A.Ghosh and S.Barman ,Prediction of Prostate Cancer cells Based on Principal Component Analysis Technique , Elsevier, International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013
[8] G.N.Satapathi, Dr. P. Srihari, Senior Member, *IEEE, Ch.* Aruna Jyothi, *S. Lavanya*, Prediction of Cancer Cell Using DSP Techniques,IEEE, International conference on Communication and Signal Processing, April 3-5, 2013, India
[9] Li, W. & Kaneko, K: "Long range correlation and partial 1/fspectrum in a non-coding DNA sequence". Europhys. Lett.,Vol.17, No.7, pp.655-660, January 1992
[10] M.Roy, S.Biswas and S.Barman (Mandal): " Identification andanalysis of coding and non-regions of a DNA sequence by Positional Frequency Distribution of Nucleotides (PFDN)algorithm". International Conference on Computers and Devices for Communication (CODEC-09), December 14-15, 2009
[11] Damanpreet Singh and J.S. Dhillon, Design of optimal IIR digital filter using Teaching-Learning based optimization technique, WSEAS Transactions on Advances in Engineering Education, vol.12, Article 2, pp. 9-18, ISSN: 1790-1979, 2015
[12]Damanpreet Singh and J.S. Dhillon Teaching Learning Based Optimization Algorithm for the Optimal Design of Higher Order BP and BS IIR Digital Filter. International Journal of Applied Engineering Research (IJAER). 10(3), pp.7727-42,2015