REVIEW PAPER ON CANCER CELL DETECTION USING MATHEMATICAL ALGORITHM

Vatsal Shah, Prof. G.B. Rathod, Neel Macwan

•

BirlaVishvakarmaMahavidyalaya
shahvatshubh@gmail.com

Abstract

Digital Signal Processing (DSP) plays an important role in the study of genes and genomics in today's time. DSP is applied to the DNA sequences and then by applying the various DSP algorithms like DFT, (Discrete Fourier Transform), DWT (Discrete Wavelet Transform). We can predict the coding region of DNA and also find out abnormality present in the coding region. This algorithm is used to find out the cancer cells. As cancer is caused due to some mutation or abnormalities present in the DNA. So this algorithm is tested in on the several cancer gene databases and normal gene databases present in the Gene bank and NCBI website that provides satisfactory results.

Keywords: Coding region, DFT Power Spectrum DWT, DSP algorithm, DNA sequences, Genomic Signal Processing

I. Introduction

Cancer is one of the most dangerous diseases in the world it is caused due to genetic mutation in the DNA.DNA means Deoxyribonucleic acid. DNA consist of information that is essential for an organism to grow and reproduce. It is made up of molecules that are known as nucleotides. Each nucleotide consists of the Phosphate group, A Sugar group, and Nitrogen base. The four types of nitrogen bases present in it are adenine(A), thymine(T), guanine(G), cytosine(C). Fig1 shows the structure of the DNA.
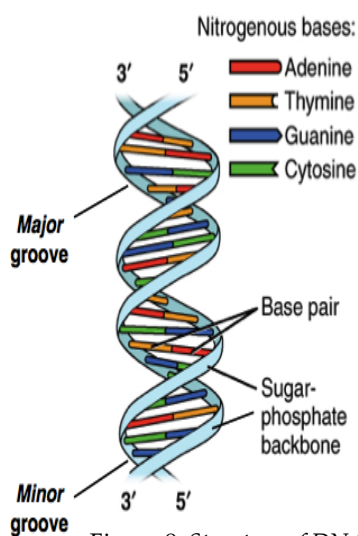

Figure 9: *Structure of DNA*

Genes are the sequences of the nucleotides present in the DNA. Organisms get the genes to inherit from the parents. Mutation in the genes causes cancer in the organism. Figure2 clearly shows the difference between normal genes and mutated cancer genes. Mutation in the genes means it is a permanent change in the sequence of the DNA which affects the chromosomes of the organism which causes several diseases these mutations can be inherited from the parent genes.
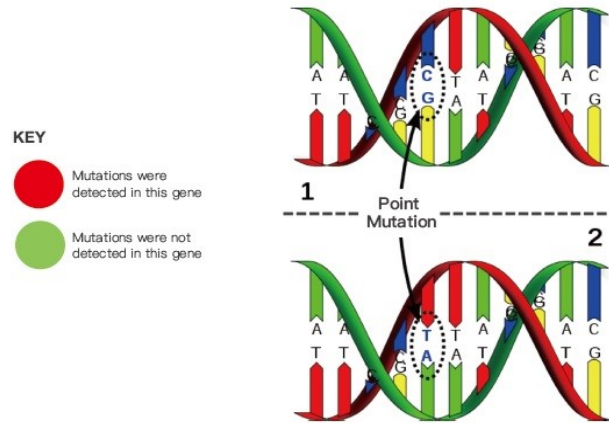

Figure 10: *Genetic Mutation*

Genomic signal processing(GSP) helps to find with the genetic mutation in the coding region of the DNA which is useful to predict the cancerous cell. Genomic signal processing deals with the study of genes database by converting the DNA sequence to the signal which helps to study the cancer cell easier other than the traditional or conventional method. It helps the mathematician, scientist, and researcher to study any kind of disease by mapping the DNA sequence properly. Further, the algorithms like Discrete Fourier transform and Discrete wavelet transform helps them to find out several Properties of the DNA like Coding region. These help them to do mathematical modeling of cells and analysis of the cell. So this helps the no biologist to study and to predict several diseases by these algorithms. So this paper deals with the mathematical modeling of the cancer cell to predict the cancer disease. The methods that are proposed in this paper is the efficient one rather than a traditional or conventional method.

II. VARIOUS CANCER DETECTION TECHNIQUES - A SURVEY

Prediction of Cancer cells using the DSP technique was proposed by G.N. Satapathi, Dr.P. Srihari, Senior, Aruna Jyothi, S. Lavanya [1]. In this paper, the authors have described the DFT algorithm on the coding region of the DNA to predict the cancerous cell. DFT means Discrete Fourier Transform it means to apply the Discrete-Time Fourier transform at the particular sample of the Signal so here the author has, first of all, mapped the bases of the DNA (A, C, G, T) to the binary it means the A and T is converted to binary value 0 and C and G is converted to binary sequence 1. Then after they have applied the DFT to the binary sequence. The formula of DFT is:

$$X_s[k] = \sum X_s[n] e^{-2\Pi kn/N}$$

(1)

for k=0,1,2, …. N-1 and n=0,1, 2, N-1.

Then the author has also applied the Power spectrum density to the sequence the formula of PSD is:

$$P_s[k]=\sum X_s[k]^2$$

(2)

Then by plotting the PSD (Power Spectrum Density) on DFT of cancer cells and normal cells they have to predict cancer cells. They have applied the IIR filter to denies the PSD signal. If the ratio of mean amplitude and mean frequency is greater than 1 then the cell is normal otherwise it is a cancer cell. DWT based Cancer Identification using EIIP was proposed by Shelli Chakraborty, Vinita Gupta. [2]. In this paper, the authors have mapped the DNA bases (A, C, G, T) to the 0.1260 0.1335, 0.0806, 0.1340. This method is known as the EIIP mapping method. Then they have applied the Discrete wavelet transform to the signal. A discrete wavelet transform is a unique algorithm to do feature extraction of the signal and to find out the statistic of the signal. The formula of the Discrete Wavelet Transform is:

$$Y[n]=(x*g)[n]=\sum x[k]\,g[n-k]$$

(3)

If the ratio of the mean amplitude of the signal and standard deviation of the signal is greater than 1 then Normal cell otherwise it is Cancer cell. Figure3 shows the Discrete wavelet analysis of the signal. And the different level of decomposition of the signal is shown in the figure. The flow chart of the algorithm is shown below:[2]
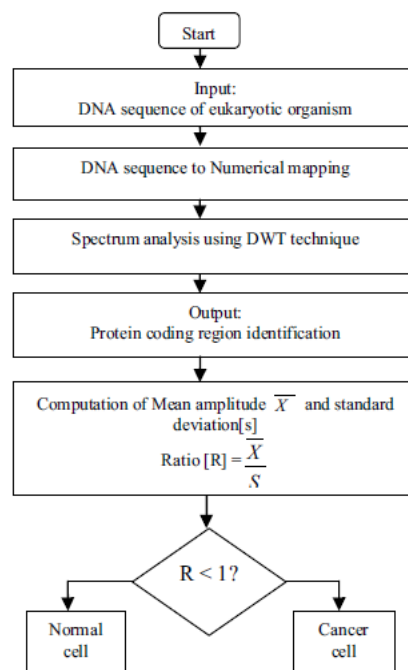

Figure 11: *Flowchart of the Research*

Detecting genetic variants of Breast cancer using different power spectrum methods was proposed by Safaa M. Naeem, Mohamed A. Eldosoky, Mai S. Mabrouk [2]. In this paper, the author has implied both the methods of mapping the DNA sequence has mentioned in both the methods they are binary mapping also known as the Voss representation method and EIIP mapping method. They have applied the method on the DNA sequence of Breast Cancer. Then they have applied Discrete Fourier transform and PSD (Power Spectrum Density) on DFT of DNA sequence of Breast Cancer. Figure 4 shows the discrete wavelet transform[2].
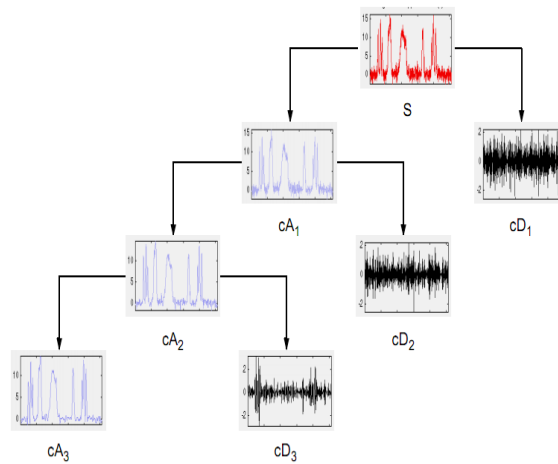
Figure.4: *Discrete Wavelet Transform*

The author has applied the other algorithm in the paper known as Welch Power Density on DFT of DNA sequence of breast cancer. Welch Power Density is used to determining the power of the signal at the different frequency is based on the concept of periodigram spectrum which includes the conversion of a signal of time domain to frequency domain. Then after they have concluded that the EIIP mapping method is best than the conventional method to predict cancer cell. Figure6 and 7 prove the same. [3]. The figure below shows the general block diagram of such kind of research:
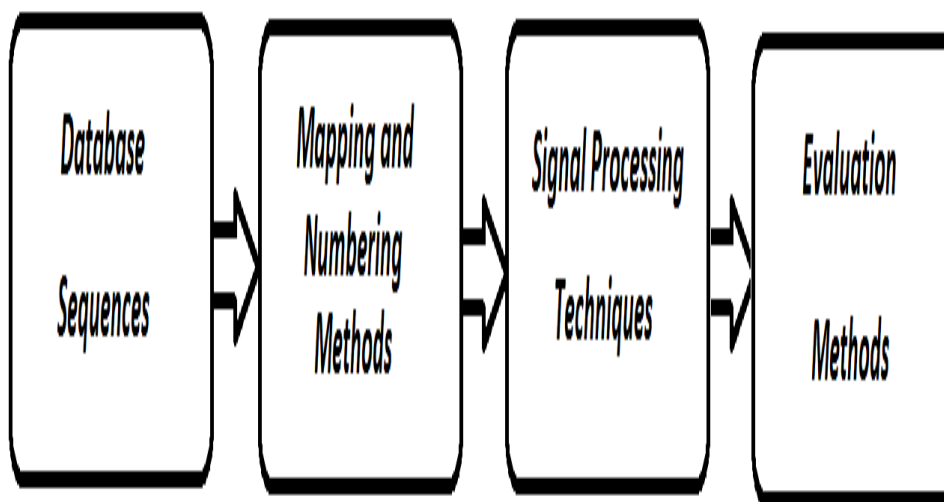

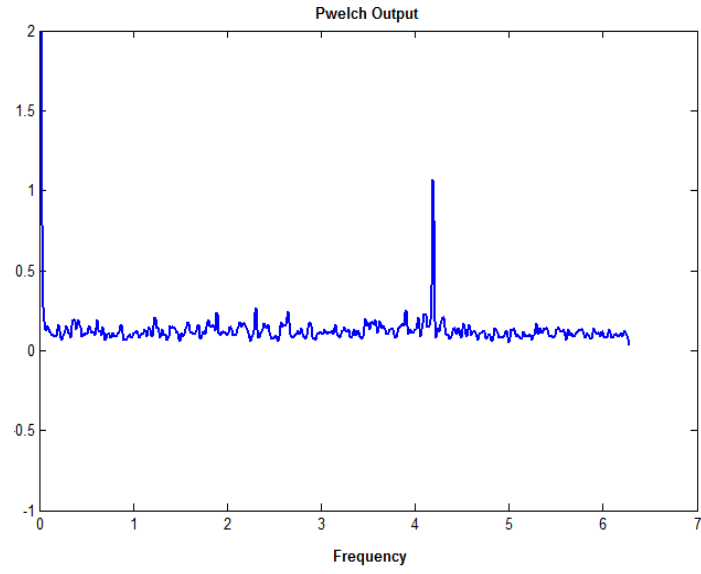Figure.5: *General block diagram of GSP*

Figure .6: *Welch power Density Spectrum of Voss representation*
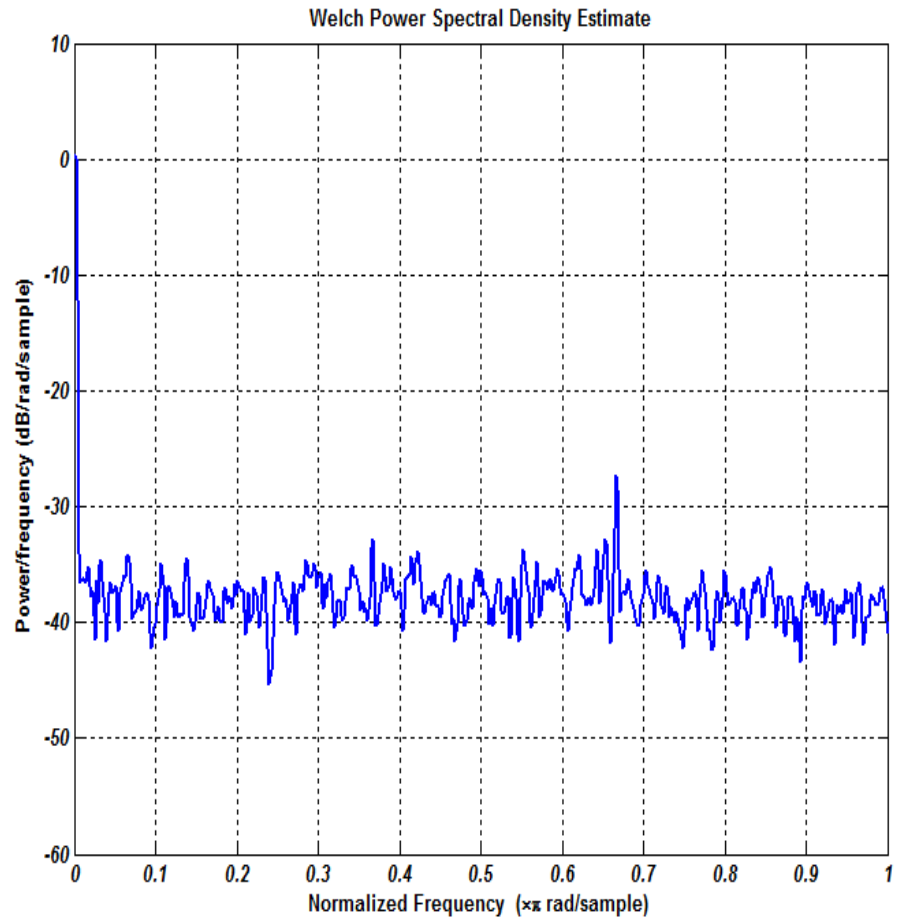

Figure. 7: *Welch Power Density Spectrum of EIIP representation*

III.   Conclusion

This paper concludes the work done to the prediction of the cancerous cell which is caused by a genetic mutation in the DNA sequence by applying the various DSP algorithms like DFT (Discrete Fourier Transform), DWT (Discrete Wavelet Transform), PSD (Power Spectral Density) and Welch Power Spectrum. The mapping of DNA by Voss representation and EIIP method. By applying these algorithms on the coding region of the DNA one can predict the cancer cell-like if the ratio of mean amplitude by its mean frequency is greater than 1 then it is normal cell otherwise cancer cell in case of DFT or if the ratio of mean amplitude by standard deviation is greater than one then it is normal cell otherwise it is a cancer cell in case of DWT. In the future, the work will be done to increase the accuracy and improving the efficiency of the plots so that cancer cells can be predicting very efficiently. These algorithms mentioned above is very efficient rather than the conventional biological approach to do prediction of the cancerous cell. The data of the cell-like DNA database is available on the NCBI website, so one should know about the name of genes, so one can easily predict the cancer cell by applying the algorithms.

References

[1]   G.N. Satapathi, P. Srihari, A. Jyothi, S. Lavanya, "Prediction of cancer cell using digital signal Processing" International Conference on Communication and Signal Processing, vol. 23, pp. 149-153, 3 April 2013.

[2]   Shilpi Chakraborty, Vinit Gupta "DWT based Cancer Identification using EIIP"2016 Second International Conference on Computational Intelligence and Communication Technology, pp.718-723.

[3]   Safaa M. Naeem, Mohamed A. Eldosoky, Mai S. Mabrouk "Detecting genetic variants of Breast cancer using different power spectrum methods",pp.147-153.

[4]   ch.mathworks.com/help/wavelet/gs/discrete-wavelet-transform.html.

[5]   Mehta, Yagnik & Dalwadi, Darshankumar. (2020). Aids Detection Using Genomics Signal Processing on DNA. 10.1007/978-981-15-4692-1_50.

[6]   B. J. Janney, S. E. Roslin and M. J. Shelcy, "A Comparative Analysis of Skin Cancer Detection based on SVM, ANN and Naive Bayes Classifier," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), Bhubaneswar, India, 2018, pp. 1694-1699, doi: 10.1109/ICRIEECE44171.2018.9008943.

[7]   M. Sajjadieh, F. Foroozan and A. Asif, "Breast cancer detection using time reversal signal processing," 2009 IEEE 13th International Multitopic Conference, Islamabad, 2009, pp. 1-5, doi: 10.1109/INMIC.2009.5383143.