

An Non-Invasive Approach of Corona Genome Detection

Kedar Patel

Electrical Engineering Dept.
Birla Vishvakarma Mahavidyalaya
Engineering College
Vallabh Vidyanagar, India
kedar1999@gmail.com

Vatsal Shah

Electronics & Communication Dept.
Birla Vishvakarma Mahavidyalaya
Engineering College
Vallabh Vidyanagar, India
shahvatshubh@gmail.com

Nisarg Patel

Electronics & Communication Dept.
Birla Vishvakarma Mahavidyalaya
Engineering College
Vallabh Vidyanagar, India
patelnisarg0711@ieee.org

Yagnik Mehta

Electronics & Communication Dept.
Birla Vishvakarma Mahavidyalaya
Engineering College
Vallabh Vidyanagar, India
yagnikmehta999@gmail.com

Abstract— Corona virus is formulated from large group of the viruses that causes illness range from common cold to more severe diseases. This virus has transmitted from animals to human beings. Millions of people throughout the world are affected to this deadly virus. So it is vital to detect this threatening virus. The Genomic Signal processing techniques is a constructive tool to detect infection as it deals with the advance research in the genetics. In this article we delineate the Genomic sequences of Corona Virus infected samples and convert the genomic data into digital signal and on applying Discrete Wavelet Transform and Fast Fourier Transform we performed mathematical modeling of the Corona-Virus gene so that it becomes effortless to predict the Corona-Virus Gene statistically. The motive of this research is to provide predicted results to drug designer for better response to this life threatening Virus. We have tested these algorithms on most significant genes of the Corona Virus as well as Normal Cells of Homosapiens. The dataset of all the genomes are available on the National Center of Biotechnology Information (NCBI) website.

Keywords—Corona Virus, Binary Mapping, Mathematical Modeling, Genomic Signal Processing.

I. INTRODUCTION

Genomic Signal Processing can be defined as the analysis, processing, and use of genomic signals to gain biological information, and the translation of that information into the system-based application that can be used to diagnose and treat genetic diseases [1] GPS aims to combine the mathematics with biology which means algorithms and techniques of signal processing are joined with the information of the genomes and process the digital genomes. Its application is: feature selection, error estimation, prediction of various genomic disease non-invasively, Classification of diseases as well as Clustering. The two important signal processing techniques that make GPS more attractive they are Fast Fourier Transform and Discrete Wavelet Transform. The structural characterization of DNA is playing a vital role in GSP. The GSP is very useful for understanding structure of DNA.

The Fast Fourier Transform is a wonderful signal processing algorithm that is used to compute N point Fourier Transform very efficiently. These algorithm is extensively use in audio processing, radar, sonar, designing of the digital filters and in

communication applications. The formula of the Fast Fourier Transform can be describing as:

$$x(k) = \sum_{n=0}^{(N/2)-1} [x(n) + (-1)^k x(n + \frac{N}{2})] W_N^{kn} \quad .1.1$$

where N stands the number of points in FFT & is known as twiddle factor.

A wavelet is a mathematical that is used to divide the continuous or discrete time signals into various scaled version of that signal. A wavelet transform is the representation of various wavelet functions. The Discrete wavelet algorithm is used to provide frequency as well as location information The formula of Haar Wavelet transform can be describe as:

$$y_n = H_n x_n \dots 1.2$$

Where: H_n is a Haar Matrix
 y_n is an output
 x_n is an input signal

Coronaviruses is formulated of group of viruses that targets the respiratory tract of homosapiens. A Coronavirus was first discovered from an infectious bronchitis virus which was present in birds. Which has capability to seriously destroy the poultry. These virus is responsible for 30 percent of common colds. The Human coronaviruses (HCov) were first discovered in the 1960s in the noses of the patients with cold The first discovered corona virus is OC43 and 229E and 2019 n-COV. The name of this viruses is because of the crown-like projections on their surfaces. This infection occurs mostly in the spring and winter months. Its effect is cold or flu which normally occurs after two or four days of infection of Coronavirus. The effects include: sneezing, a runny nose, fatigue, a cough, fever, a sore throat, exacerbated asthma. This virus spreads through coughing and sneezing without covering the mouth, by physical contact with person who is already infected by coronavirus, contact with a surface that posses virus through noise mouth or any organ. The crown structure of coronavirus is shown in the figure below:

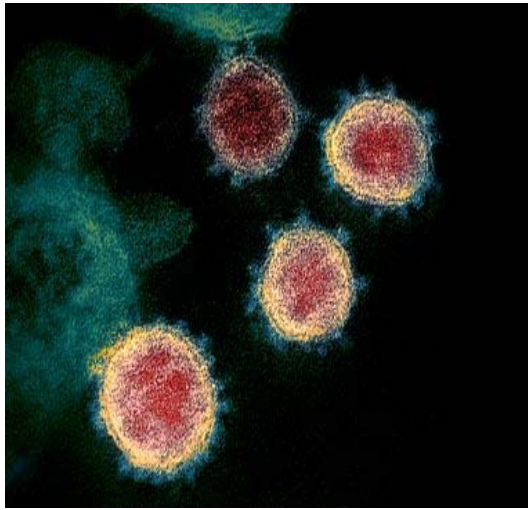


Figure. 1. Structure of Coronaviruses [3]

The NCBI Stands for National Center for Biotechnology Information. It posses information regarding genes genomic data & biomedical information. The genomic data is this website not limited up to humans but all organisms. So any user can find the data regarding any Genes.

II. BINARY MAPPING OF DNA SEQUENCE

The living organism has a most essential component that helps them to grow is Deoxyribonucleic Acid. As we known that DNA is formed by four chemical elements they are Adenine, Cytosine, Guanine, Thymine. In DNA sequence only these 4 elements are present one can easily convert these signal to binary. Hence binary mapping means to depict these elements into the binary one. Here Adenine is assigned as binary **00** (decimal **0**), Cytosine is assigned as binary **01** (decimal **1**), Guanine is assigned as binary **10** (decimal **2**), Thymine is assigned as binary **11** (decimal **3**). After applying these mapping technique the DNA sequence will convert to binary signal and it becomes easy to apply signal processing algorithm for prediction. Figure below shows the binary mapping concept. It is the eminent component of our research.

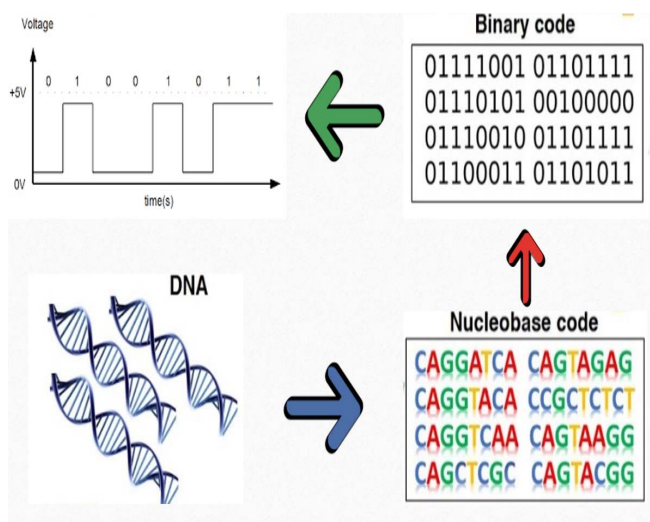


Figure. 2. Binary Mapping of DNA sequence [2]

III. OUR RESEARCH METHODOLOGY

The methodology we have adopted in the prediction of the Corona Virus Gene is explained below:

- We have taken 60,000 samples of Normal Genome as well as RNA converted to DNA sequence of the Coronavirus genome from the database that is available on the NCBI website.
- Then map these genome sequences as discuss in the section II.
- We have first of all determine the Fast Fourier Transform on the converted genomic sequence (binary mapped genome). In order to perform signal analysis
- Then Determine the Discrete Wavelet Transform on the resultant genomic Sequence to compute statistics of the Signal.
- Based on the results of the Discrete wavelet transform. Compute ratio of Standard deviation (S) by Change in Mean Amplitude ($\Delta\mu_x$) of the binary DNA sequence. We have concluded to result that:

- If $\frac{S}{\Delta\mu_x}$ is not range [1 1.77] then it is Genome of Coronavirus.
- If $\frac{S}{\Delta\mu_x}$ is in range [1 1.77] then it is Genome of Normal Cells.

IV. FLOWCHART

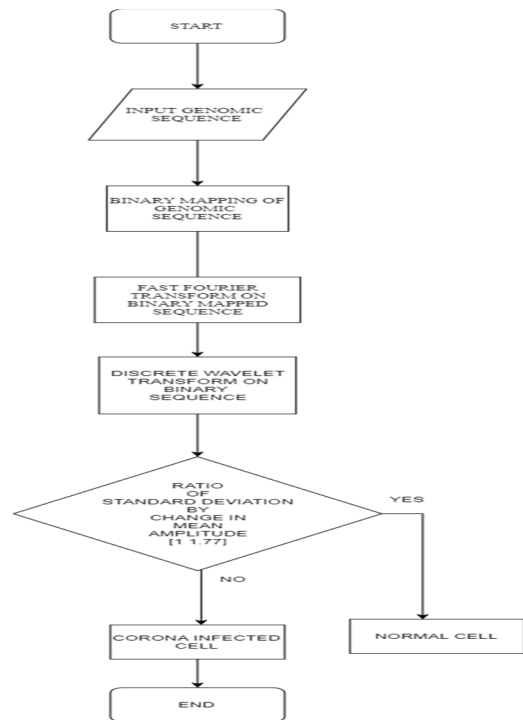


Figure. 3. Flow chart of the Research

V. FAST FOURIER ANALYSIS OF THE NORMAL GENOME & CORONA GENOME

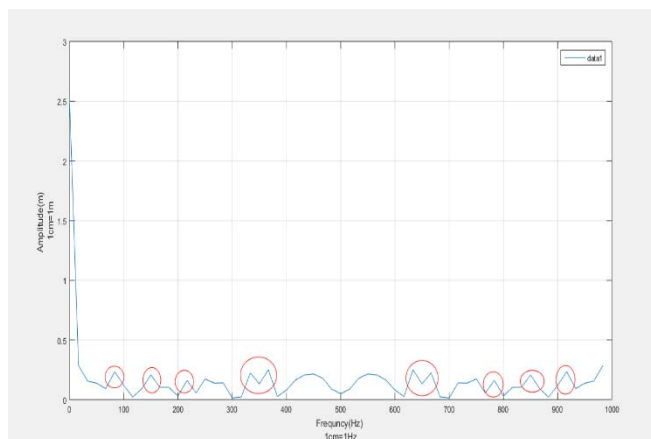


Figure. 4. N Coronavirus Gene

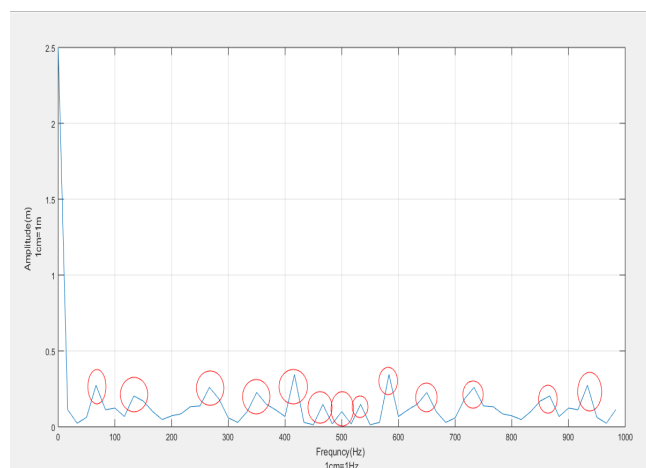


Figure. 7: HBG1 Normal Gene

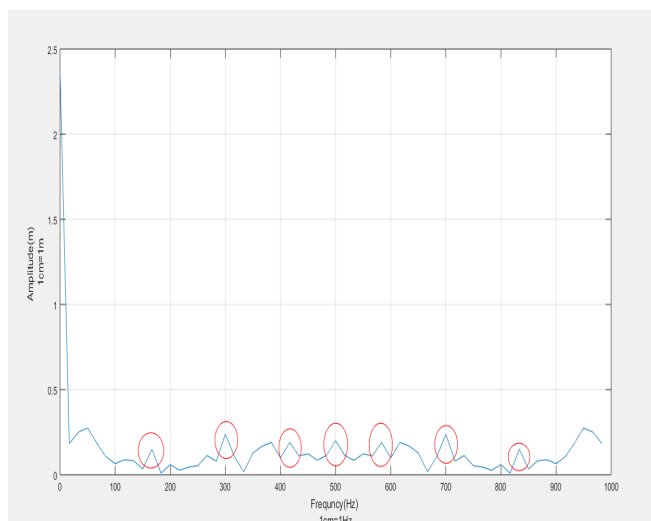


Figure.5. HBB1 Normal Gene

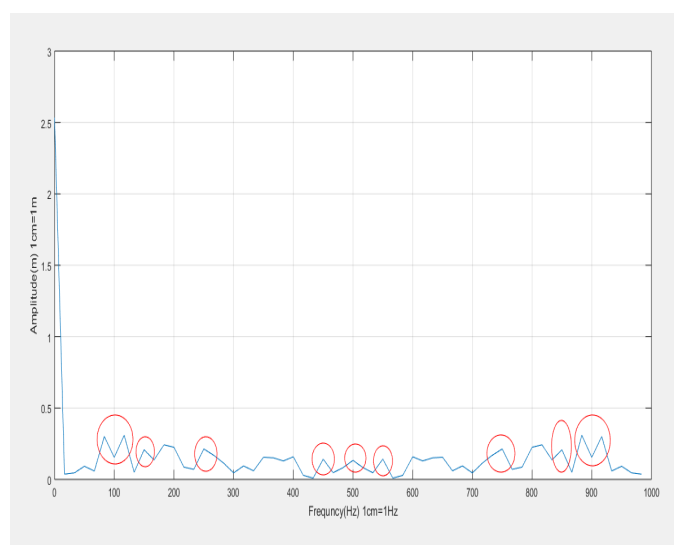


Figure. 8: 229E Coronavirus Gene

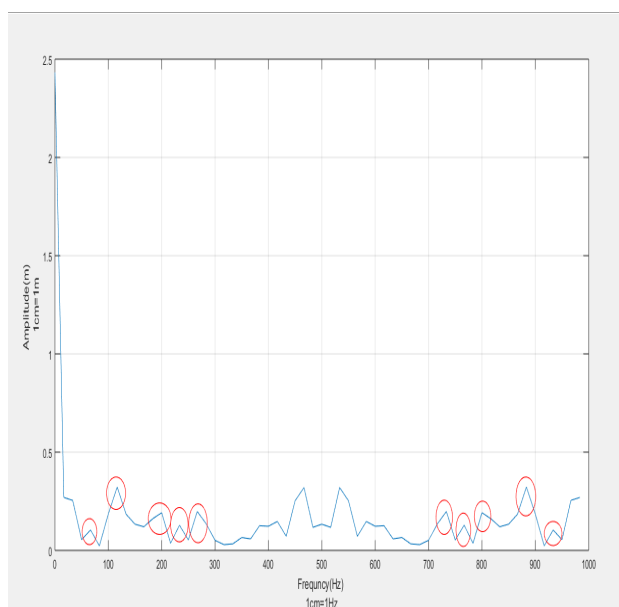


Figure. 6. N-COV Coronavirus Gene

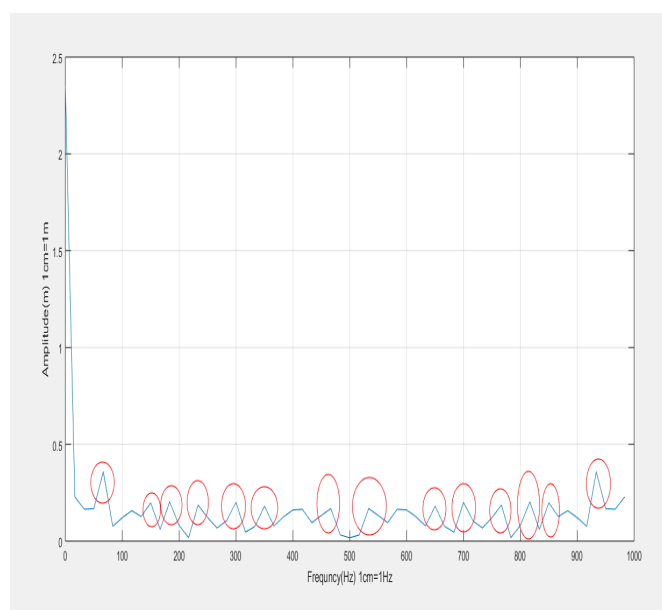


Figure. 9. Globin Normal Gene

VI. DISCRETE WAVELET OF THE NORMAL GENOME & CORONA GENOME

Table.1: Coronavirus Genome

SR No:	Genome	Standard Deviation (S)	Change in Mean Amplitude of Signal ($\Delta\mu_X$)	$\frac{S}{\Delta\mu_X}$
1	229E	0.9472	1.083	0.874
2	Beta Coronavirus	0.8634	0.367	2.3525
3	N	0.8181	0.9	0.9098
4	N-COV	0.9273	1.103	0.84070
5	ORF1B	0.9204	0.433	2.215
6	SARS-NCOV	0.8452	1.108	0.7628

Table.2: Normal Genome

SR No:	Genome	Standard Deviation (S)	Change in Mean Amplitude of Signal ($\Delta\mu_X$)	$\frac{S}{\Delta\mu_X}$
1	HBB1	0.7933	0.45	1.76288
2	HBB2	0.7811	0.45	1.7357
3	HBB3	0.922	0.9	1.0244
4	HBB4	0.7912	0.7920	1.0034
5	GATA	0.7901	0.7900	1.0012
6	GLOBIN	0.7973	0.45	1.7717

VII. RESULTS & DISCUSSION

The plots of the Fourier Transform of binary converted DNA sequences of Normal as well as corona genome as well as Discrete wavelet analysis of the binary signal are as shown in the above sections. So one can interpret from the Fourier plots that corona genome has more pulse width compare to the normal genome moreover there are twin peaks found in the corona virus genome and it is not present in the normal genome. There are higher amplitude peaks in the plots of normal genome compare to the corona virus genome. While performing discrete wavelet transform on these binary sequence, we have found from the result of the statistics of the signal that if the ratio of standard deviation by change in mean amplitude of binary converted genome is in range [1 1.77] then the predicted genome is of normal. Otherwise the predicted genome is of corona-virus infected. so based on results one can easily understand the characteristics of the genome.

VIII. CONCLUSION

In these article there is a description of effective algorithm to predict the Corona Virus Gene. And essential part of the research is the delineation of the Genomic sequences the delineation technique that is used here is Binary Delineation. Then Fast Fourier analysis as well as Discrete Wavelet technique has been applied to genomic data for detection Corona Virus Genome. So in the future, it has a great scope in early detection of Corona Virus. We experimentally found that the plots of the Fast Fourier Transform of the Genome of the Corona Virus and Normal Genome is found to be somewhat similar that's why by applying Discrete Wavelet Transform we done Statistical Analysis of the Corona Virus Genome so in Future this algorithm has a great scope of predicting the Corona Virus Genome non-invasively. It fortunately helps the drug developer to develop the antinode to cure the person whose genome is attack by Corona Virus. So these is how one relate mathematics with biology for the prediction of the various deadly diseases.

REFERENCES

- [1] Hong-Qian Wang, Hau-San Wong, De-Shaung Huang, Jun Shu." Extracting gene regulation information for cancer classification", Pattern Recognition,2007.
- [2] <https://ashutoshviramgama.com/dna-data-storage-synthetic-dna-future-of-storage/>
- [3] <https://www.nih.gov/news-events/nih-research-matters/novel-coronavirus-structure-reveals-targets-vaccines-treatments>
- [4] <https://www.gizchina.com/scientists-announce-high-resolution-crystal-structure-the-new-coronavirus-3cl-hydrolase/>
- [5] <https://www.ncbi.nlm.nih.gov/>
- [6] <https://www.medicalnewstoday.com/articles/256521.php>
- [7] <https://www.mathworks.com/>