

# Cell-Free DNA Pattern Identification and its Classification using Convolution Neural Network

Nithya.P, Arthy.P.S,Prabhu Kumar.S

**Abstract** – Circulating cell DNA (cfDNA) design identification assumes a cardinal job in fetal drug, transplantation and oncology. Be that as it may, it has additionally demonstrated to be a biomarker for different maladies. There are numerous order strategies by which the acknowledgment and arrangement should be possible. So as to have a superior time unpredictability and improve the precision further, this strategy targets distinguishing and arranging the general DNA examples and ailments related with them utilizing cfDNA Images in a Convolution Neural Network. A probabilistic method is used for cfDNA image feature extraction, fragmentation and interpretation. Graphical User Interface is the platform where this method is employed since it uses visual indicators in place of text-based interface. The aftereffects of this test demonstrate that the Convolution Neural Network calculation can perceive cfDNA successions accurately and successfully with no dubiety. Prepared with examples, the CNN can effectively characterize the picture surrendered to coordinated and unparalleled examples with numerical highlights.

**Keywords** – cfDNA, Convolution Neural Network, Graphical User Interface, GLCM feature extraction, Bioinformatics.

## I. INTRODUCTION

DNA pattern recognition is crucial in bioinformatics. Modern biology and biochemistry make acute use of DNA technology in genetic engineering, where a number of methods have been developed to obtain DNA from organisms via blood, semen, urine, skin which are used for medical research and forensic purposes

.Circulating cell-free DNA (cfDNA) are DNA fragments released into the blood stream in case of casualties such as stroke, trauma, diabetes. It can be seen during the course of pregnancy. Hence, it plays a major role in fetal medicine. It is used to check the compatibility between donor and receptor during transplantation. The cfDNA residuals are found in patients with cancer and genetic disorders.

Here, we portray the utilization of our cfDNA design acknowledgment and arrangement utilizing Convolution Neural Network. Our foundation comprises of two sections:

**Revised Manuscript Received on November 08, 2019.**

**Nithya.P**, Department of Electronics and Communication Engineering, Vel Tech Multi Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Chennai, India

**Arthy.P.S**, Department of Electronics and Communication Engineering, Vel Tech Multi Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Chennai, India

**Prabhu Kumar.S**, Department of Electronics and Communication Engineering, Vel Tech Multi Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Chennai, India

First, include extraction utilizing Gray Level Co-event Matrix, which shows the highlights in factual arrangement. Second, Convolution Neural Network is trained to analyze the features obtained from the input image, compare them with the ideal data set features and classify them accordingly as Pattern Matched or Pattern Unmatched. If the pattern is matched, then the person is deemed healthy and if unmatched, he is affected.

## II. PROPOSED SYSTEM

### Introduction

Circulating cell- free DNA (cfDNA) plays a vital role in fetal medicine and in organ transplantation. It is an extensive tool in the technological development of bioinformatics that includes trauma, sepsis, myocardial infarction, sickle cell diseases, stroke and more.

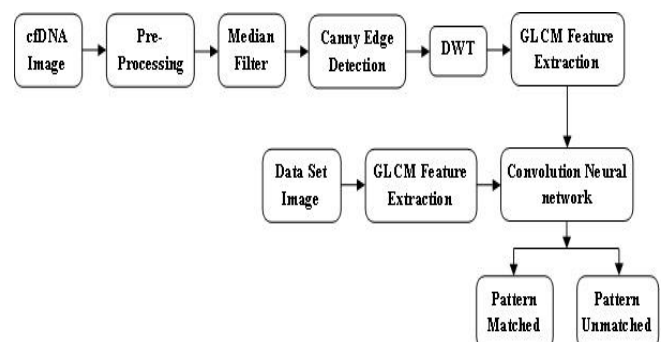
The aim is to identify and classify the DNA patterns using cfDNA images. It focuses to design and develop a cfDNA pattern recognition and classification system using Convolution Neural Network.

In order to robustly extract the precise details of the cfDNA in terms of numerical values, we employ GLCM to obtain maximum features from the cfDNA pattern. This method extracts the features such as contrast, correlation, energy, homogeneity.

The edges of the cfDNA are detected using Level 4 decomposition of Discrete Wavelet Transform. Canny Edge detection provides the visual edges from which the data is extracted. And then, using Convolution

Neural Network which has better performance on time complexity, we analyze and classify them into matched and unmatched cfDNA patterns.

### B. FUNCTIONAL BLOCKDIAGRAM



**Fig 1.cfDNA pattern recognition and classification system**

## C. WORKING OF cfDNA PATTERN RECOGNITION AND CLASSIFICATION

The block diagram given shows the explained procedure involved in cfDNA pattern recognition and classification method. The input cfDNA image obtained in the form of microscope image from the collected blood or urine sample is given to the pre-processor.

The pre-processing procedure which consists of grayscale conversion and resizing of the input image in accordance with the size of the data set aims at the improvement of the image data, suppression of the undesired distortions and enhancement of some image features required for further processing.

The pre-prepared picture is bolstered into the middle channel for clamor evacuation purposes. The middle channel works over a window by choosing the middle power in the window and has a preferred position of less affectability than different channels.

**Canny Edge Detection** is utilized to speak to the edges outwardly. It works by applying Gaussian channel to smoothen the clamors in the picture. It figures the inclination size of the picture utilizing limited contrast approximations for the halfway subsidiaries. Non maxima concealment is applied to the inclination greatness to stifle the frail edges and twofold limit calculation is utilized to identify and connection edges. At long last, Hysteresis is utilized.

Edge has the most high recurrence data of an advanced picture however it very well may be obscured by commotion evacuation. So a Level-4 Wavelet Decomposition DWT is utilized in the upgrade of vertical and flat edges of the separated picture given just as in further clamor decrease. The single level deterioration is rehased in a falling way to further build the goals and the estimate co-proficient disintegrated with high and low pass channels.

**GLCM feature extraction** process is applied to the DWT image in order to extract the data from the image in a statistical format. It characterizes the texture of an image by calculating the spatial relationship between the pixels, creating a matrix and extracting numerical data from the created matrix.

The features obtained from this method are:

- Contrast – the proportion of nearby varieties in the GLCM.
- Correlation – the proportion of joint likelihood event of determined pixel sets.
- Energy – the entirety of squared components in the GLCM.
- Homogeneity – the proportion of closeness of the dissemination of components in the GLCM to the GLCM

corner to corner.

The data set provides the ideal reference cfDNA of a healthy human being. It is compared with the input in order to classify the cfDNA provided as affected and unaffected. As similar to the input process, GLCM is applied to the data set to extract the respective features.

Convolution Neural Network is the classifier used. The features from the input set and the features from the data set are compared and classified. The Neural Network is trained in such a way that a large number of inputs can be sampled. If the pattern is matched i.e., the statistical features extracted from the cfDNA provided should come close to the features extracted from the reference data set, then the input provided person is healthy. If the pattern is unmatched, then the input provided person is affected.

## D. ABBREVIATIONS AND ACRONYMS

DWT- Discrete Wavelet

Transform GLCM - Gray

Level Co- Occurrence CNN -

Convolution Neural Network

GUI- Graphical User Interface

GUIDE- Graphical User Interface Development Environment

## III. RESULT AND DISCUSSION

Thus, **Cell-free DNA pattern recognition and its classification using Convolution Neural Network**, unique cfDNA pattern recognition is designed and developed. It successfully recognizes the cfDNA patterns given and classifies them accordingly. The outputs of various stages of the proposed system are given in the section below.

Gray scale conversion reduces the storage i.e., the number of bits are reduced. Median filter proves to be more efficient in noise reduction than the other filters since it does not alter the original image.

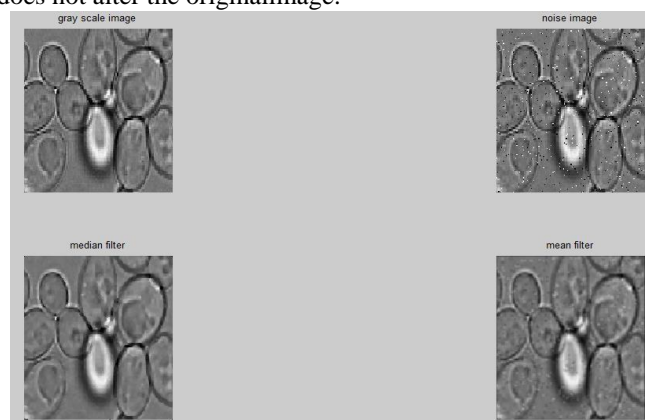


Fig 2. Pre-processed Image

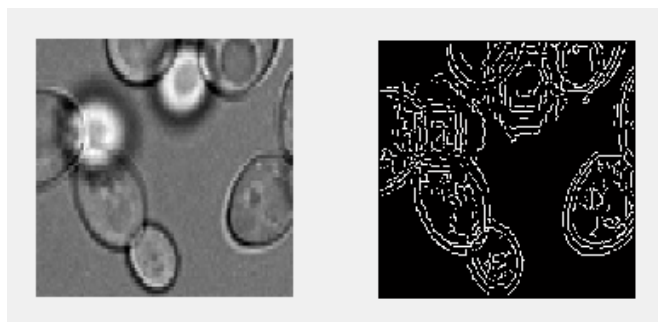


Fig 3 Canny Edge Detection

Canny edge detection uses the gradient magnitude to display the edges visually. Strong edges are enhanced and weak edges are depressed.

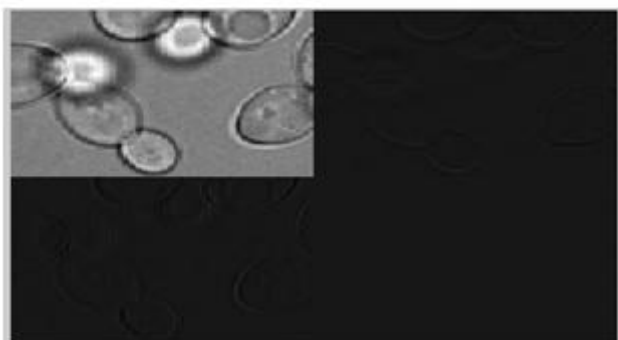


Fig 4 DWT applied cfDNA image

DWT uses non-decimated method to retrieve the edge information. Level 1 Decomposition is cascaded to get Level 4 Decomposition.

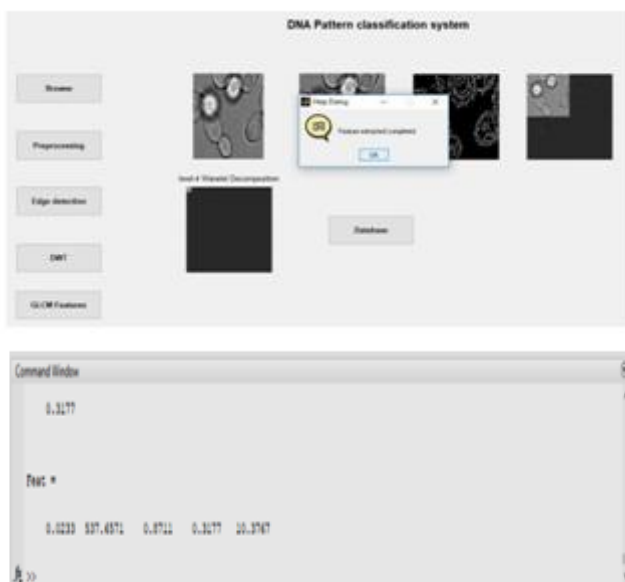


Fig 5 GLCM Feature Extraction

Energy, Contrast, Correlation, Homogeneity and Entropy are the features extracted in the above window respectively from the GLCM Matrix.



Fig 6(a). Pattern matched output



Fig 6(b). Pattern unmatched output

Fig 6(a) and Fig 6(b) shows the cfDNA classification output obtained using Convolution Neural Network in terms of Pattern Matched and Pattern Unmatched.

#### IV. CONCLUSION

Thus, a method where the DNA patterns can be recognized, analyzed and classified into healthy and affected DNA patterns is developed. It provides accurate results on whether the patterns match or not. By using Convolution Neural Network, this method can be implemented on larger scale and has improved the time complexity. Other than fetal medicine and genetics, it is mainly used for forensic purposes. There are numerous directions in which this work can be extended. First of all, it can be improved to name the accurate disease rather finding from a number of susceptible possibilities.

#### REFERENCES

1. Z. Xizhi, "The **Application** of Wavelet Transform in Digital Image Processing," 2008 International Conference on MultiMedia and Information Technology, Three Gorges, (2008), pp. 326-329. doi:10.1109/MMIT.2008.134
2. Sriram KB, Relan V, Clarke BE, Duhig EE, Windsor MN, Matar KS, et al. (2012). "Pleural fluid cell-free DNA integrity index to identify cytologically negative malignant pleural effusions including mesotheliomas". BMC Cancer 12:428.
3. Liimatainen SP, Jylhvi J, Raitanen J, Peltola JT, Hurme MA. (2013). "The concentration of cell-free DNA in focal epilepsy". Epilepsy Res 105(3):292-8.
4. Chen, Shifu & Liu, Ming & Zhang, Xiaoni & Long, Renwen & Wang, Yixing & Han, Yue & Zhang, Shiwei & Xu, Mingyan & Gu, Jia. (2017). "A Study of Cell-free DNA Fragmentation Pattern and Its Application in DNA Sample Type Classification". IEEE/ACM Transactions on Computational Biology and Bioinformatics. PP. 1-1. 10.1109/TCBB.2017.2723388.
5. Khier S, Lohan L. "Kinetics of circulating cell-free DNA for biomedical applications: critical appraisal of the literature". Future Sci OA. 2018;4(4):FSO295. Published 2018 Feb 23. doi:10.4155/fsoa-2017-014.

