DNA FRAGMENTATION & PATTERN RECOGNITION USING NEURAL NETWORKS

Gorle Shyam Kumar, Saikiran, Manekantta
Department of ECE, Bharath Institute of Higher Education & Research

Abstract - Classification and Pattern Recognition of DNA modules are rather important at various aspects. Be it detection and treatment of viral infections, genetic problems, diagnostic confirmation and multiple disorder classifications in the Medical Domains, finding abundant resource generation and Crime scene identification in analytic prospects, DNA Pattern Recognition forms a major crux. Several aspects have to be taken care whilst classifying DNA patterns and recognition. Neural Networks has been the technical epitome of some time. This paper deploys using Neural Networks for classification and pattern recognition for essential DNA Samples. MATLAB Data Processing is being used in the paper.

Keywords*: -DNA Pattern, Fragmentation, MATLAB, Dataset Processing, Neural Networks*

I. INTRODUCTION

Humans are characterized with their unique attributes. DNA residuals are some among them. The need of DNA Fragmentation and classification owing to medicinal and scientific frameworks is a very big business market and thus the advancement of affirmative innovations in the emerging field of DNA Pattern recognition is of great market and business potential. Significant applications of DNA Pattern Recognition involves bio science, law requirement and observation, identification check, criminal examinations, identification of a particular person among a bunch and much more.

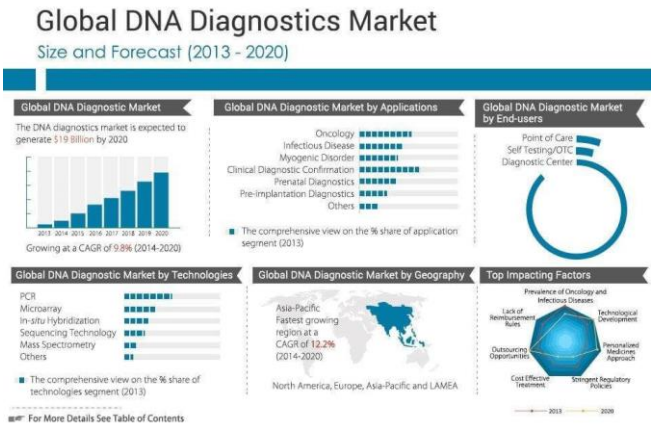


Fig 1 . DNA Pattern Recognitions market demands

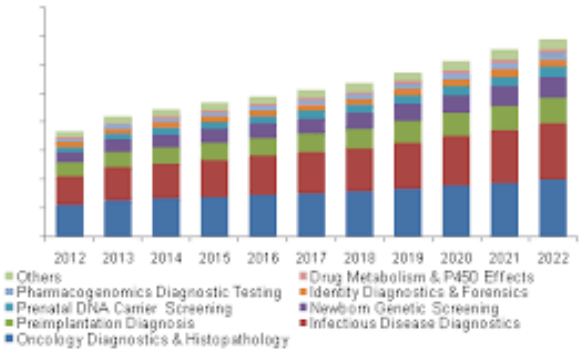


Fig 2 . DNA Analysis – Market Potential

The above figure details the applications of DNA recognition and the worldwide market potential in this field of emerging science. Unlike specialized applications, the concept of DNA Recognition has a huge potential world wide, thus illustrated from the figure below.

Starting from 130M, the potential of DNA Analysis is predicted to be closed to 1000M in the 2022

.

II. EXISTING SYSTEMS

Existing methods propose to save the images as datasets and process them using MATLAB, the Image Processing Math works tool. These methods are time consuming and not cost effective. Moreover, DNA Analysis in these methods are not variant and are not effective during various conditions. We require a effective DNA Analysis solution, that is more accurate in predictive technologies and marking systems.

Several other papers have proposed making use of various filters and DWT. These are accurate to a certain extent, but updated technologies such as Neural Networks can prove more accurate results, which are the need of the hour. This level of accuracy is essential for minor medical applications, which on the other side may prove to have adverse results as they may prompt to be used for diagnostic and operational benefits.

Existing DNA Segmentation Algorithms take Inputs, segmentation of the same is done using specific algorithms, features of the same are extracted and it is then classified as per the comparision results with the existing datasets.

This can even sometimes prove wrong or inadequate as this is totally relied on the existing datasets and the decision is made based only on the available datasets, which may prove to be wrong. Moreover updates on the datasets are not provided as in this case.

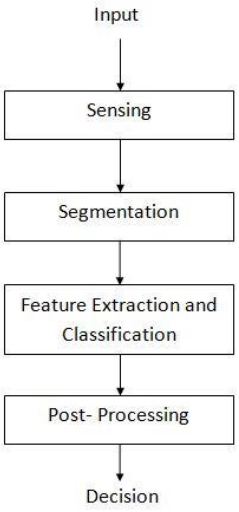


Fig 3 . Existing Methods of DNA Pattern Recognition

III. PROPOSED METHODOLOGY

We propose to inculcate neural networks for manipulating, analysis and detect pattern of DNA that can be used for diagnostic purposes. The process flow of the DNA Segmentation and Pattern Identification consists of several stages. Preprocessing, Filtering, Edge Detection, Feature Extraction, Processing, Comparison & prediction.

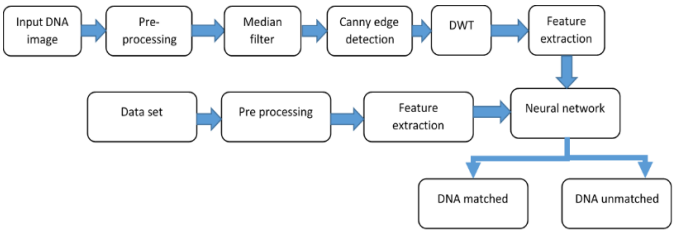


Fig 4 . Block Diagram of the Proposed Method of DNA Classification & Recognition

A. Database Creation :

The Process of capturing and storing images of the DNA Segemented samples form the first step of the DNA Pattern recognition. Datasets are updated from time to time and the featured , compared datasets are again added to the datasets to form a complete reliable datasource for comparative analysis.

B. Pre Processing:

Pre Processing form the major crux of the implementation of the project. Though conventional methods are being used, it is the algorithm that makes it different from the existing papers.

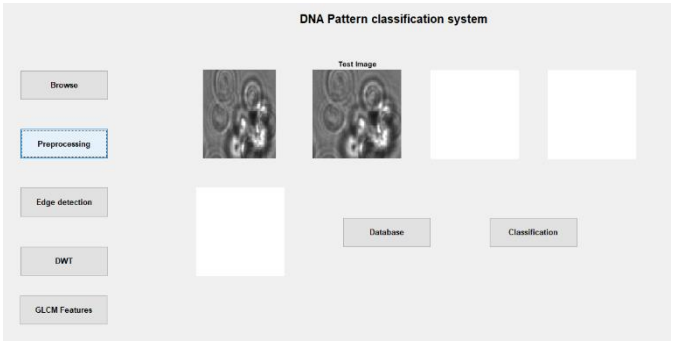


Fig 5. Screenshot of the Pre Processing Process

Any input image cannot be processed or fed as input to any other filters. It has to be preprocessed, ie converted to binary formats that is within the permissible limits and can be processed a multiple speeds and comparatively accurate without disturbing the features of the same. This is done in this segment. Filtering is done to remove noise and permit the same as per the reliable frequency levels in terms of signal processing.

B. *Edge Detection*

Edge detection is done to the input image to make it comparable to the dataset, so that the image processed is done comparatively without any hassles or loss. Smoothing, blurring of the image to remove noise, finding gradients, Non-maximum suppression, Double thresholding, Edge tracking by hysteresis are all a part of the Edge detection and Canny Edge detection is deployed in our project.

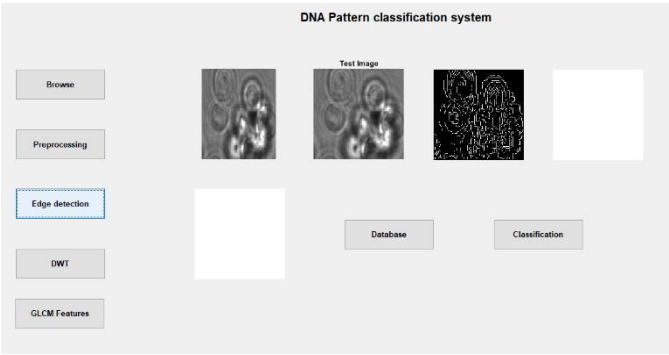


Fig 6. Screen Shot of the Edge Detection Process Block

C.DWT & Feature Extraction

Discrete Wavelet parameters, reduces the previously continuous basis set of wavelets to a discrete and orthogonal / orthonormal set of basis wavelets.

$\psi m,n(t) = 2m/2\ \psi(2mt - n)$; m, n $\in$Z such that - $\infty <$ m, n $< \infty$

The 1-D DWT is given as the inner product of the signal x(t) being transformed with each of the discrete basis functions.

$$Wm,n = < x(t), \psi m,n(t) > ; m, n \in Z$$

The 1-D inverse DWT is given as:

$$x(t) = \sum_m \sum_n W_{m,n} \psi_{m,n}(t) \quad ; m, n \in Z$$

The processed image with the applied filters , DWT with the feature extracted , we get a comparable image so as we can apply the same to the NN Algorithm , so as to compare and make the decision.

D.NEURAL NETWORK APPLICATIONAL LAYER

Gray level Co-occurrence Matrix Features is deployed in the paper along with the k Means Clustering Algorithm. The basic structure of the NN Classifier is in fig 7.

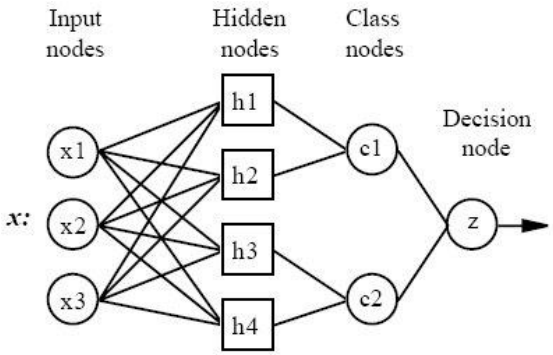


Fig .7.Architecture of the NN Classifier

As depicted the input nodes x1 , x2 and x3 are processed simultaneously with the coherent hidden nodes and then classified as class nodes. The class nodes are basically the pre processed input image and the dataset to which the input image is compared to . Algorithmic matches lead to Z – the decision node . This is the comparative decision of the NN Algorithm the implies that the DNA is matching to or not matching to the existing pattern and hence results arrived at.

The NN Algorithm implications are done at nodal levels comparatively than that of the image levels in the existing systems, which make this paper unique, reliable and more accurate.

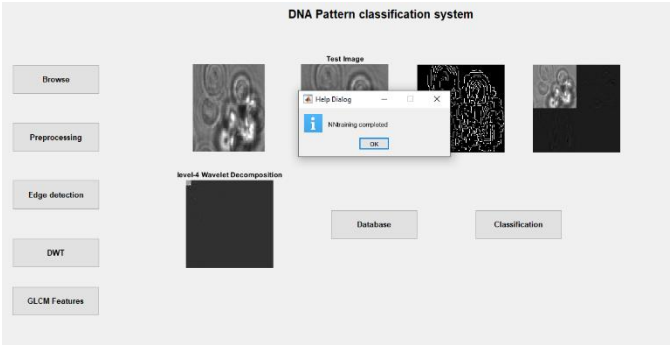


Fig .8.Architecture of the NN Classifier

IV. DATA INPUTS & PROCESSING

Input data and existing datasets are a major criterion in the paper. The existing datasets are obtained by the strenuous updates from the medicinal updates and the same can also be used as the input data for vertificational purpose

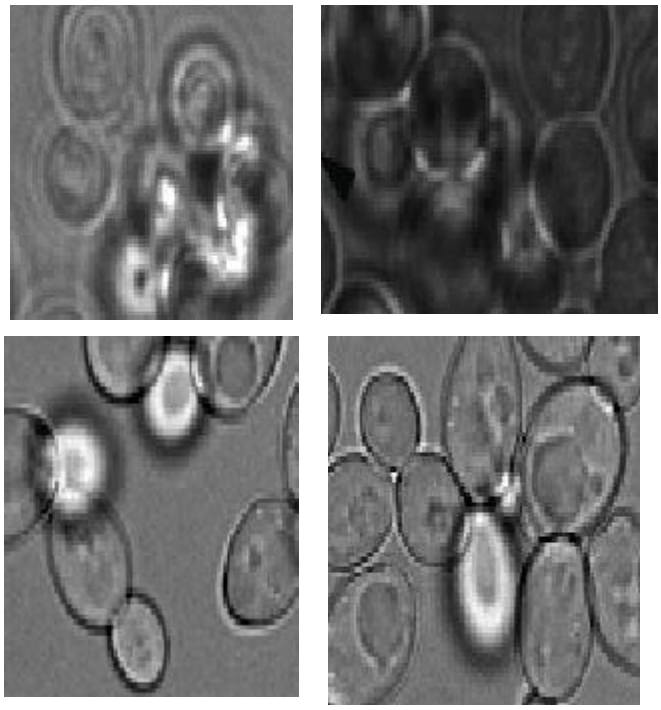


Fig 9 .Performance Diagram of Proposed Method of Face Attendance marking

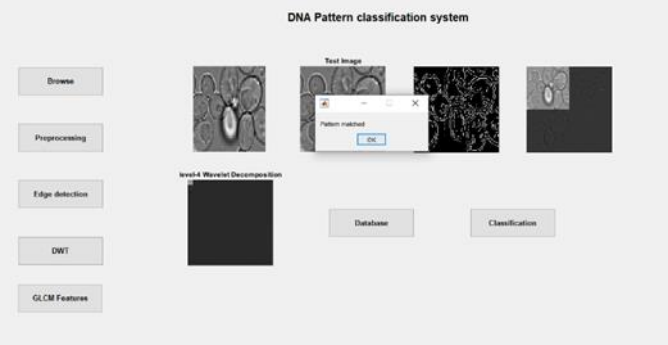
V. SOFTWARE SYSTEM REQUIREMENTS

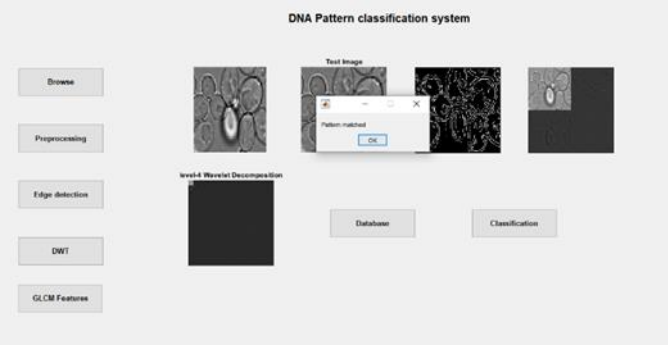
A. *MATLAB:*

MATLAB has been used in the paper. Matlab is extensively used for Image processing and classification applications. It is an extensive tool specifically deployed for Image and Signal processing. The MATLAB tools allows to provide us with the Input, allow to write preprocessing algorithms, compare with the existing datasets and apply NN Algorithms between different multiple images and processes.

VI. RESULTS AND FUTURE ENHANCEMENTS

As per the requirement, the datasets are preloaded in the database and any of the comparative image is used as the input image for verification purpose. Up to 1000 images can be compared based on the speed and capacity of the system.
Image represents the criteria wherein the dataset does not match with the existing dataset. This indicates that there is no pattern recognized in the existing samples of the DNA. Next is the comparative analysis of the DNA Sample , with the match found, so that the DNA Pattern is recognized and is available for diagnostic application.





The paper is accurate and reliable in terms of the Algorithms which is being deployed. Future Enhancements of the paper shall comprise of deploying different algorithms that may get updated from time to time and also comparing with the large amount of datasets. Henceforth as the time consumption is also a factor to be considered, it has to be noted that the accuracy is the major factor that has to be considered in medical and crime applications.

VII. REFERENCES

[1] Mandel P, Metais P. (1948). Les acides nucleiques du plasma sanguin chez l'homme [in French]. C R Seances Soc Biol Fil 142:241-243.

[2] otezatu I, Serdyuk O, Potapova G, Shelepov V, Alechina R, Molyaka Y, Anan'ev V, Bazin I, Garin A, Narimanov M, Melkonyan H, Umansky S, Lichtenstein AV. (2000). Genetic analysis of DNA excreted in urine: a new approach for detecting specific genomic DNA sequences from cells dying in an organism. Clin Chem 46:1078-1084.

[3] 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.

[4] Liimatainen SP, Jylhv J, Raitanen J, Peltola JT, Hurme MA. (2013). The concentration of cell-free DNA in focal epilepsy. Epilepsy Res 105(3):292-8

[5] Stroun M, Lyautey J, Lederrey C, Olson-Sand A, Anker P. (2001). About the possible origin and mechanism of circulating DNA: Apoptosis and active DNA release. Clin Chim Acta 313(1-2):139- 42.

[6] Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, et al (2001). DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res 61(4):1659-65

[7] Chiu RWK, Chan KCA, Gao Y, Lau VYM, Zheng W, et al. (2008). Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc Natl Acad Sci U S A 105: 20458-20463