International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 04 |Apr 2019 www.irjet.net p-ISSN: 2395-0072

DNA Fragmentation Pattern and Its Application in DNA Sample Type Classification

R.Ashwin[1],M.Junaid Hasan[2],S.Dominic Rosario[3],S.koushik[4],Mr.Vijay Ananth[5]

[1,2,3,4] UG Student, Department of Electronics and Communication Engineering, SRM Institutte of Science and Technology, Tamil Nadu, India

[5]Assistant Professor (O.G), Department of Electronics and communication Engineering, SRM Institutte of Science and Technology, Tamil Nadu, India

---

Abstract—*DNA pattern recognition is a key problem in a*

bioinformatics and biomedical informatics. We solve this problem by use of the probability method and metric instead of traditional frequency metric. And then, we put forward our neural network which has better performance on time complexity than some sequence alignment algorithms in the same field. The results of the contrast experiments show that our Neural Network algorithm can recognize DNA sequences correctly and effectively without any ambiguities. Trained with patterns, the network successfully classified image given as input.

Key Words: Neural Network algorithm, Bioinformatics

I. INTRODUCTION

THE SIZE of DNA fragments gives important information for the construction of physical genome maps and genotyping. Inparticular ,by knowing the DNA fragment length and the DNA molecule profile. It is possible to investigate the properties of single DNA molecules or of DNA-protein interactions. It is possible, for example, to distinguish between different DNA secondary structures and also to establish if and in what manner a ligand binds to DNA . For DNA Fragmentation K-means method are very common. Their limitations are the low speed (processing times of 2 hours or more) and large amount of DNA samples required for analysis. This avoids errors introduced from operator bias and increases the amount of information available for further analysis such as DNA intrinsic curvature and dynamic structure analysis, critical for the understanding of several key biological processes (e.g., DNA packaging, transcription, replication, recombination, repair, and nucleosome stability and positioning

II.LITERATURE SURVEY

The component extraction is done in Pattern Recognition based DNA sequence compressor in 2012 technique used by Pattern Recognition based DNA sequence compression.

Anil K Jain et.al, in this paper of pattern recognition gave an overview of this field as the primary goal of pattern recognition is supervised or unsupervised classification.

In this papers in which pattern recognition has been formulated, the statistical approach has been most intensively studied and used in practice. Nowadays, neural network techniques and methods imported from statistical learning theory have been increasing attention.

Scott C Newton et.al, most real data structures encountered in speech and image recognition and in medical and many other decision making tasks are quite complex in nature and rather difficult to organize for programing autonomous and optimal control and recognition systems. This paper presents a modular, unsupervised neural network architecture which can be used for clustering and classification of complex data sets. The adaptive fuzzy leader clustering architecture is a hybrid neural-fuzzy system which learns on-line in a stable and efficient manner research topics and applications which are at the exciting and more challenging field in DNA.

III.PROPOSED FRAMEWORK

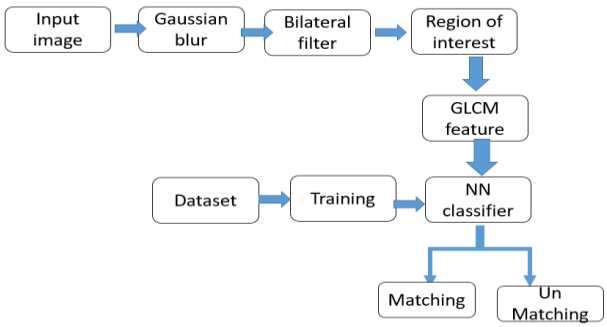

Fig. 1. DNA Fragmentation Pattern

III. GAUSSIAN BLURRING

Gaussian Blurring is used to reduce image noise and reduce image quality.The Gaussian function is used in numerous research areas.It defines a probability distribution for noise or data.It is a smoothing operator in image processing. Gaussian smoothing is commonly used in edge detection. In edge-detection algorithms are sensitive to noise.The 2-D Laplacian filter, built from a discretization of the Laplace operator, is highly sensitive to noisy environment. Using a Gaussian Blur before edge detection aims to reduce the level of noise in the image, which improves the result of the following edge-detection algorithm.

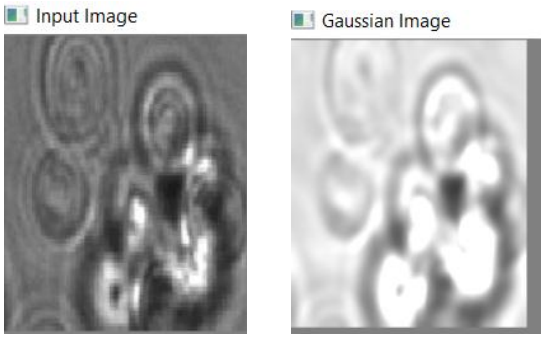

Fig .2.Gaussian Blurring

IV.BILATERAL FILTER

A bilateral filter is an edge-preserving, and noise-reducing smoothing filter for images. It replaces the intensity of each pixel with a weighted average of intensity values from nearby pixels from given images. The weight can be based on a Gaussian distribution. The weights depend not only on Euclidean distance of pixels, but also on the radiometric differences such as range differences, such as color intensity, depth distance, etc. This preserves sharp edges in images The bilateral filter that was introduced as an edge preserving selective smoothing mechanism.

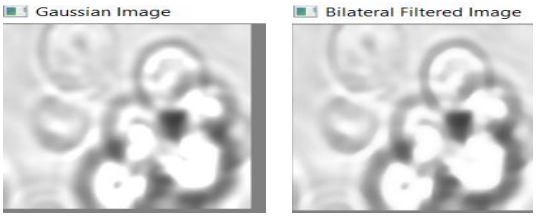

Fig.3.Bilateral Filter

V.REGION OF INTEREST

A region of interest (often abbreviated ROI). The concept of a ROI is commonly used in many application areas such as medical imaging, the boundaries of a tumor may be defined on an image, for the purpose of measuring its size. After the Bilateral filter,It allows the detection of the ROI through the matchmaking of a complete specification of the exact shape of the image. This sub-module comprises the following steps are load the image, create on the step and use it as input image. The pixels that belong outside of the ROI images are set to 1 and pixels outside the ROI are set to 0.

VI.GLCM PROPERTIES

The (GLCM) gray-level co-occurrence matrix , also known as the gray-level spatial dependence matrix. To create GLCM, use the gray co-matrix function. The gray co-matrix function creates a gray-level co-occurrence matrix by calculating how often a pixel with the intensity value i occurs in a specific spatial relationship to a pixel with the value j. The spatial relationship of GLCM is defined as the pixel of interest and the pixel to its immediate right but you can specify other spatial relationships between the two pixels.

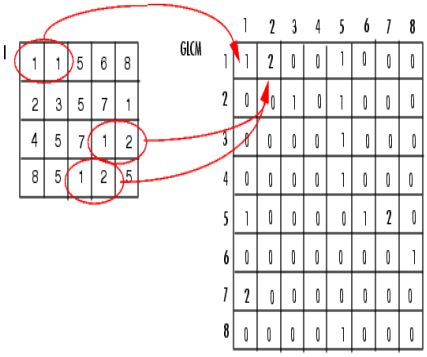

Fig.4.GLCM

VII.GLCM FEATURES

Contrast is used to measures the local variations in the gray-level co-occurrence matrix. Coorelation is used in GLCM matrix and used to measure measures the joint probability occurrence of the specified pixel pairs .Energy gives the sum of squared elements in the GLCM and also known as uniformity or the angular second moment.Homogenity is used to measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

Sl. No	Texture Features	
	Feature	*Formula*
1	Contrast	$\sum_{i,j} \lvert i-j \rvert^2 p(i,j)$
2	Correlation	$\sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j}$
3	Energy	$\sum_{i,j} p(i,j)^2$
4	Entropy	$-\sum_{i,j}(p(i,j).*\log(p(i,j))$
5	Homogeneity	$\sum_{i,j} \frac{p(i,j)}{1+\lvert i-j \rvert}$

Fig.5.GLCM FEATURES

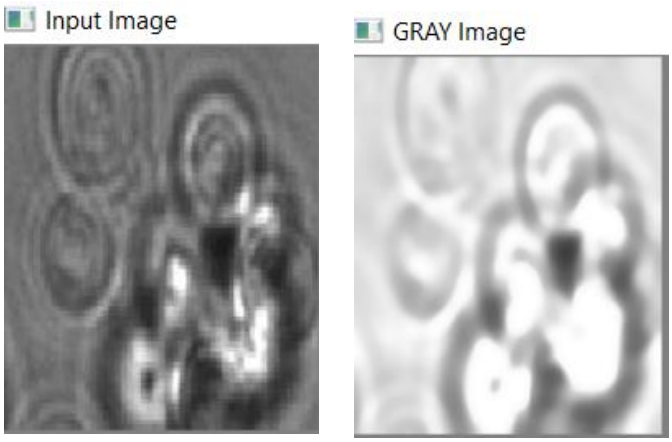

Fig.6.Gray Images

VIII.NN CLASSIFIER

k-NN (k-nearest neighbors algorithm) is a non-parametric method, where the function is only approximated locally. The k-NN algorithm is the easiest of all machine learning algorithms for classification, It is a useful technique can be used to assign weight to the contributions of the neighbors, and the nearer neighbors contribute more to the average than the more distant ones. A common weighting scheme consists in giving each neighbor a weight of 1/d, where d is the distance to the neighbour in the given sample. The neighbors are taken from a set of objects for for k-NN classification or the object property value for k-NN regression . The K-NN can be used for the algorithm, though no explicit training step is required. The k-NN algorithm is that it is sensitive to the local structure of the data.

IX.IMAGE SEGMENTATION

The image segmentation is the process of partitioning a digital image into multiple segments such as sets of pixels, also known as super-pixels. The aim of Image segmentation is to simplify and change the representation of an image into something that is more meaningful and easier to analyze the image Image segmentation is typically used to locate objects and boundaries like lines, curves in images. Image segmentation is the process of giving a label to every pixel in an image such that pixels with the same label share certain characteristics. The image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image .Each of the pixels in a region are similar with some characteristic or computed property, such as color, intensity and texture. Adjacent regions are significantly different with respect to the same characteristics.

Fig.7. Segmented Region & Extracted Image

X.CONCULSION

We used this tool to test the DNA samples outside the training and validating datasets, and it gave near 100% accuracy for classifying hundreds of samples .Deep analysis of different chromosomes gave similar patterns for autosomes and sex chromosomes. Comparing to the patterns of whole sequence, the patterns of DNA were found closer than supposed, suggesting that all blood cells may share same mechanisms to produce DNA. The DNA fragmentation patterns of long or short DNA fragments also have similar patterns, which indicates the short DNA fragments are mainly from the release from cells, not produced by the degradation of longer fragments when they are circulating. The DNA has more than 98% for humans is non-coding, meaning that these sections do not serve as patterns for protein sequences.

XI.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.