Journal of Bioinformatics and Sequence Analysis Vol. 3(6), pp. 100-115, September 2011
Available online at http://www.academicjournals.org/JBSA
ISSN 2141-2464 ©2011 Academic Journals

Review

Identification of unique repeated patterns, location of mutation in DNA finger printing using artificial intelligence technique

B. Mukunthan[1]*, N. Nagaveni[2] and A. Pushpalatha[3]

[1]Department of Master of Computer Applications and Research, SVS College of Engineering, Coimbatore, Tamil Nadu, India- 614030.
[2]Department of Mathematics and Research, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India.
[3]Department of Mathematics and Informatics, Government Arts College, Udumalpet, Tamil Nadu, India-642126.

Accepted 17 June, 2011

The proposed Neural-Fuzzy pattern recognition (NFPR) system discussed in this paper effectively reduces the complication in precisely analyzing and interpreting human deoxyribonucleic acid (DNA) sample. In this novel approach, the perfect blend made of bioinformatics and a competitive method of neural networks technique, which has the advantage over conventional computation technique, in their ability to solve problem that do not have an algorithmic solution or the available solutions are also too complex to be found, results in efficient DNA pattern analysis algorithm that identifies repeated patterns in the given human DNA sample assisting in generation of unique identification number of an individual, location of occurrence of mutation in the mutated DNA sample with utmost prediction accuracy.

Key words: Neural-Fuzzy resonance mapping, competitive learning, NFPR processor, Input generator, preprocessor, discriminator, DNA profiling, DNA sequence, FASTA format.

INTRODUCTION

The genome (Joe and John, 1999) is the entirety of an organism's hereditary information which is encoded either in deoxyribonucleic acid (DNA) or, for many types of virus, in ribonucleic acid (RNA). The role of DNA sequences has become indispensable for many biological researches. DNA sequencing is applied in various fields such as diagnostic, biotechnology, forensic biology and biological systematic.

The DNA sequences of thousands of organisms have been decoded and stored in databases. A comparison of genes within a species or between different species can show similarities between protein functions, or relations between species. With the growing amount of data, it became impractical to analyze DNA sequences manually. A pattern (Richard et al., 2006; Donald, 2005) is essentially an arrangement or an ordering, in which some organization of underlying structure can be said to exist, that is, a pattern can be referred to as a quantitative or structural description of an object or some item of interest. A set of patterns that share some common properties can be regarded as pattern class (Phipps, 1996) in our case the unique repeated nucleotide sequence from the given human DNA sample.

Neural networks (Advances in Neural Networks issn, 2006) can process information in parallel, at high speed, and in a distributed manner. Neural networks which are simplified models of the biological neuron system, made up of highly interconnected neural computing elements that have the ability to learn and thereby acquire knowledge and make it available for use. Neural networks are capable of learning by examples to solve

*Corresponding author. E-mail: mukunth_bmk@yahoo.co.in, mukunthan.bmk2011@gmail.com . Tel: +918870353613.

unknown or untrained instances of the problem, if it's aptly trained.

Neural networks architectures (Stephen et al., 2009; Robert, 2007) can be trained with known examples of a problem before they are tested for their inference. They can, therefore, identify new objects previously untrained. Neural networks (Carpenter and Grossberg, 1987) are robust systems and are fault tolerant. They can therefore, recall full patterns from incomplete, partial or noisy patterns. Network architectures (John et al., 2008) have been classified into various types based on their learning mechanisms and other features.

In Competitive Learning method those neurons which respond strongly to input stimuli have their weights updated, when an input pattern is presented, all neurons in the layer compete and the winning neuron undergoes weight adjustment. Hence it is a "Winner-takes-all" strategy.

Neural networks (Advances in Neural Networks issn, 2006) suitable particularly for pattern classification problems in realistic environment is Neural-Fuzzy resonance mapping (NFRM) (Carpenter and Grossberg, 2010), it is a vast simplification of fuzzy resonance mapping which possess reduced computational overhead and architectural redundancy.

DNA PROFILING AND SEQUENCING

DNA profiling (Stephen and David, 2003) also called DNA testing, DNA typing, or genetic fingerprinting, is a technique employed by forensic (Norah and Keith, 2002; Joe and John, 1999; John and Brent, 2005) scientists to assist in the identification of individuals on the basis of their respective DNA profiles. DNA profiles (David, 2004) are encrypted sets of numbers that reflect a person's DNA makeup (David, 2008; Andreas, 2001), which can also be used as the person's identifier. DNA sequencing theory addresses physical processes related to sequencing DNA .The term DNA sequencing refers to sequencing methods for determining the order of the nucleotide bases—adenine, guanine, cytosine, thymine and uracil (rare case) in a molecule of DNA. Single nucleotide poly-orphisms (Computational Intelligence and Bio inspired Systems, 2005) are a DNA sequence variation occurring when a single nucleotide A, T, C, or G in the genome (Julie, 2001) (or other shared sequence) differs between members of a species (Des and willie, 2000) (or between paired chromosomes in an individual).

For example, two sequenced DNA fragments from different individuals (Michael, 2007), AAGCCTA to AAGCTTA, contain a difference in a single nucleotide. Various DNA sequence formats available are: Plain sequence format, EMBL format, GCG format, GCG-RSF (rich sequence format, Gen Bank format, IG format and given sample is used as an input to neural-fuzzy pattern recognition (NFPR) processor which is to be interpreted and analysed. Fuzzy representation of nucleotide bases in NFPR processor is Adenine (A)-0.1, Thymine (T)-0.2, Guanine (G)-0.3, Cytosine (C)-0.4, Uracil (U)-0.5.The concept of clustering logical and illogical sequence is shown in Figure 1.

NEURAL-FUZZY PATTERN RECOGNITION PROCESSOR

Learning input generator

The input generator is used for input normalization and it represents the presence of particular feature in the input patterns and its absence. Various cases for generating normalized learning inputs are shown in Table 1.

Learning inputs

$$\mathrm{LIN}_{i,\,n} = I_1, I_2 \ldots, I_p \qquad (1)$$

Where $0.1 \le i \le 0.5, 0.1 \le n \le 0.5$

and $p = 4$ (sizeof learning input, weights of NFPR processor)

Various cases for learning input normalization is given in Table 2.

Activation function generator

When coded input patterns from input generator are presented to NFPR-Processor all output nodes become active to varying degrees. The output activation denoted by activation function (ACFj) for the j^{th} output node. Where LIN is the learning input and LIW_j is the corresponding learning input weights.

Activation unction

$$\mathrm{ACFj} = \frac{\left|\mathrm{LIN} \ \Lambda \ \mathrm{LI\ Wj}\right|}{\alpha + \left|\mathrm{LI\ Wj}\right|} \qquad (2)$$

Here α is kept as a small value close to 0 it's about 0.0000001.The node which registers the highest activation function is deemed Winner node, that is:

$$\mathrm{Winnernode} = \max(\mathrm{ACFj}) \qquad (3)$$

In the event of more than one node emerging as the winner owing to the same activation function value some FASTA format. A sequence file in FASTA format of a

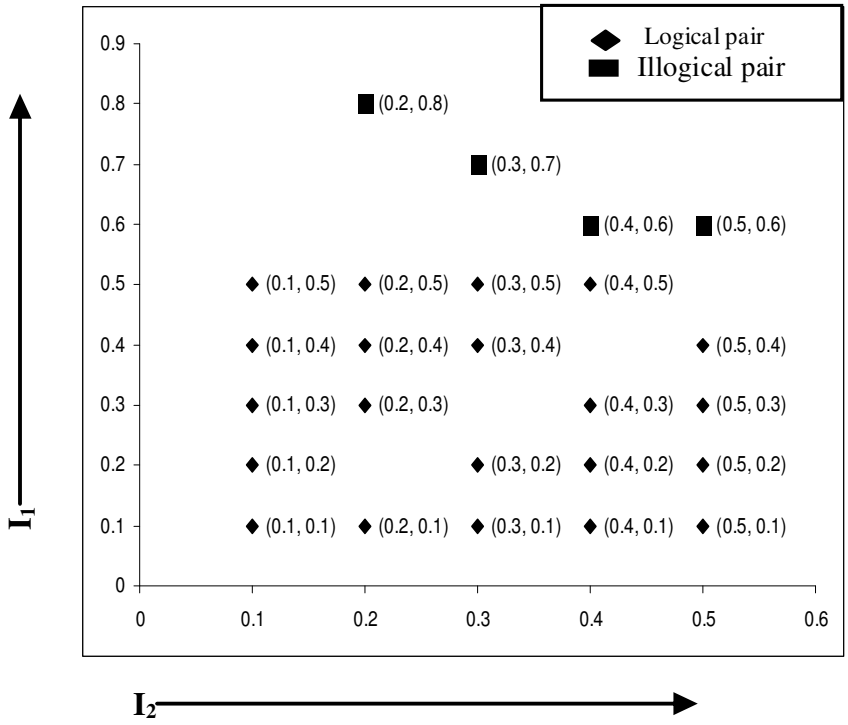

Figure 1. Concept of Clustering Logical and Illogical sequence.

mechanism such as choosing a node with the smallest index may be devised to break the tie.

Match function generator

The match function (MAF) which helps to determine whether the network must adjust its learning parameters is given by:

Matching function

$$\mathrm{MAFj} = \frac{\left| \mathrm{LIN} \wedge \mathrm{LI\ Wj} \right|}{\left| \mathrm{LIN} \right|} \qquad (4)$$

The MAF association with the vigilance parameter (ρ) decides on whether a particular output node is good enough to encode a given input pattern or whether a new output node should be opened to encode the same.

The network is said to be in a state of resonance, if the match function value exceeds vigilance parameter. However, for a node to exhibit resonance, it is essential that it not only encodes the given input pattern but should also represent the same category as that of the input pattern. The network is said to be in state of mismatch reset if the vigilance parameter exceeds match function, Such a state only means that the particular output node is not fit enough to learn the given input pattern and thereby cannot update its weights even though the category of the output node may be the same as that of the input pattern. This is so, since the output node has fallen short of the expected encoding granularity indicated by the vigilance parameter. If match function is greater than vigilance parameter and category of input pattern is not same with the learning input, the vigilance parameter is updated and is given by:

$$\rho = \mathrm{MAF} + \delta \qquad (\delta = 0.001) \qquad (5)$$

The weight updating equation of an output node *j* when it proceeds to learn the given input pattern *LIN* is given by:

Weight for Inference (WFI)

$$\mathrm{WFIj}^{\mathrm{new}} = \beta(\mathrm{LIN} \wedge \mathrm{WFIj}^{\mathrm{old}}) + (1-\beta)\mathrm{WFIj}^{\mathrm{old}} \qquad (6)$$

$$\text{where} \quad 0 \le \beta \le 1 \qquad (\beta = 1)$$

The computation involved in generating WFI and category for Inference (CFI) for some nucleotide pairs is

Table 1. Generating weights for inference (WFI), category for inference (CFI) from learning inputs.

Nucleotide pair		**A,A***	**A,U***	**T,A***	**T,T****	**T,U***	**G,A***	**G,G****	**G,U***	**C,A***	**C,C****	**C,U***	**U,A***	**U,U*****
Category		**L**	**L**	**L**	**ILL**	**L**	**L**	**ILL**	**L**	**L**	**ILL**	**L**	**L**	**ILL**
Learning input(LIN)		0.1,0.1, 0.9,0.9	0.1,0.5, 0.9,0.5	0.2,0.1, 0.8,0.9	0.2,0.8, 0.8,0.2	0.2,0.5, 0.8,0.5	0.3,0.1, 0.7,0.9	0.3,0.7, 0.7,0.3	0.3,0.5, 0.7,0.5	0.4,0.1, 0.6,0.9	0.4,0.6, 0.6,0.4	0.4,0.5, 0.6,0.5	0.5,0.1, 0.5,0.9	0.5,0.6, 0.5,0.4
ρ		0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5, 0.600+ δ	0.601	0.601	0.601
Learning input weights	LIW(1)	=LIN	0.1,0.1, 0.9,0.9	0.1,0.1, 0.9,0.5	0.1,0.1, 0.8,0.5	0.1,0.1, 0.8,0.5	0.1,0.1, 0.8,0.5	0.1,0.1, 0.7,0.5	0.1,0.1, 0.7,0.5	0.1,0.1, 0.7,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5
	LIW(2)	~	~	~	=LIN	0.2,0.8, 0.8,0.2	0.2,0.8, 0.8,0.2	0.2,0.8, 0.8,0.2	0.2,0.7, 0.7,0.2	0.2,0.7, 0.7,0.2	0.2,0.7, 0.7,0.2	0.2,0.6, 0.6,0.2	0.2,0.6, 0.6,0.2	0.2,0.6, 0.6,0.2
	LIW(3)	~	~	~	~	~	~	~	~	~	~	~	~	0.5,0.1, 0.5,0.9
Activation function	ACF(1)	0.9999	0.7999	0.9375	0.7999	0.9999	0.9333	0.8751	0.9999	0.9285	0.9230	0.9999	0.9230	0.8461
	ACF(2)	~	~	~	~	0.8499	0.5999	0.8999	0.8888	0.6111	0.8888	0.9375	0.6249	0.9375
	ACF(3)	~	~	~	~	~	~	~	~	~	~	~	~	0.7499
Highest Activation Function		ACF(1)	ACF(1)	ACF(1)	ACF(2)	ACF(1)	ACF(1)	ACF(2)	ACF(1)	ACF(1)	ACF(2)	ACF(1)	ACF(1), ACF(2)	ACF(2)
Category Match / Mismatch		Match	Match	Match	Match	Match	Match	Match	Match	Match	Mismatch Match	Match	Match	Match
Match Function	MAF(1)	1.0000	0.8000	0.7500	0.6000	0.7500	0.7000	0.6000	0.7000	0.6500	0.6000	0.6500	0.6000	0.5500
	MAF(2)	~	~	~	~	0.8500	0.6000	0.9000	0.8000	0.5500	0.8000	0.7500	0.5000	0.7500
	MAF(3)	~	~	~	~	~	~	~	~	~	~	~	~	0.7500
Weight for inference and category for inference	WFI(1)	0.1,0.1, 0.9,0.9	0.1,0.1, 0.9,0.5	0.1,0.1, 0.8,0.5	0.1,0.1, 0.8,0.5	0.1,0.1, 0.8,0.5	0.1,0.1, 0.7,0.5	0.1,0.1, 0.7,0.5	0.1,0.1, 0.7,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	**0.1,0.1, 0.6,0.5**
	CFI(1)	L	L	L	L	L	L	L	L	L	L	L	L	**L**
	WFI(2)	~	~	~	0.2,0.8, 0.8,0.2	0.2,0.8, 0.8,0.2	0.2,0.8, 0.8,0.2	0.2,0.7, 0.7,0.2	0.2,0.7, 0.7,0.2	0.2,0.7, 0.7,0.2	0.2,0.6, 0.6,0.2	0.2,0.6, 0.6,0.2	0.2,0.6, 0.6,0.2	**0.2,0.6, 0.5,0.2**

Table 1. Contd.

CFI(2)	~	~	~	ILL	ILL	ILL	ILL	ILL	ILL	ILL	ILL	ILL	**ILL**
WFI(3)	~	~	~	~	~	~	~	~	~	~	~	0.5,0.1, 0.5,0.9	**0.5,0.1, 0.5,0.9**
CFI(3)	~	~	~	~	~	~	~	~	~	~	~	L	**L**

ρ=VIGILANCE PARAMETER ,*- CASE1,**-CASE2,***CASE3.

given below and shown in Table 1:

Delta (δ) =0.001; Beta (β) =1; and Alpha (α) =0.0000001,

For nucleotide pair (AA)

Category to be trained: L (Logical)

Fuzzy Equivalent (FE): 0.1(A), 0.1(A)

Learning Input (LIN) =0.1, 0.1, 1-0.1, 1-0.1
= 0.1, 0.1, 0.9, 0.9 (Case1)

Rho (ρ) =0.5, $\mathrm{LIW(1)= AI = 0.1,0.1,0.9,0.9}$

$$\mathrm{ACF(1)} = \frac{\left|0.1,0.1,0.\ \ 9,0.9\ \ \Lambda\ 0.1,0.1,0.\ \ 9,0.9\right|}{0.0000001\ +\left|0.1,0.1,0.\ \ 9,0.9\right|}$$

$$= \frac{\left|0.1,0.1,0.\ \ 9,0.9\right|}{2.0000001} = \frac{\left|2.0\right|}{2.0000001} = 0.9999$$

$$\mathrm{MAF(1)} = \frac{\left|0.1,0.1,0.\ \ 9,0.9\ \ \Lambda\ 0.1,0.1,0.\ \ 9,0.9\right|}{\left|0.1,0.1,0.\ \ 9,0.9\right|}$$

$$= \frac{\left|0.1,0.1,0.\ \ 9,0.9\right|}{2.0} = \frac{\left|2.0\right|}{2.0} = 1.0000$$

As it's the first learning input,

$$\mathrm{WFI(1)} = 0.1,0.1,0.9,0.9\ ,\ \mathrm{CFI(1)} = \mathrm{L}\ .$$

For nucleotide pair (AU)

Category to be trained: L (Logical).
Fuzzy Equivalent: 0.1(A), 0.5(U).

Learning Input (LIN) = 0.1, 0.5, 1-0.1, 1-0.5
=0.1, 0.5, 0.9, 0.5 (Case1)
Rho (ρ) =0.5,

$$\mathrm{LIN} = 0.1,0.5,0.9,0.5\ ,\ \mathrm{LIW(1)} = 0.1,0.1,0.9,0.9$$

$$\mathrm{ACF(1)} = \frac{\left|0.1,0.5,0.\ \ 9,0.5\ \ \Lambda\ 0.1,0.1,0.\ \ 9,0.9\right|}{0.0000001\ +\left|0.1,0.1,0.\ \ 9,0.9\right|}$$

$$= \frac{\left|0.1,0.1,0.\ \ 9,0.\ 5\right|}{2.0000001} = \frac{\left|1.6\right|}{2.0000001} = 0.7999$$

$$\mathrm{ACF(1)} = \frac{\left|0.1,0.5,0.\ \ 9,0.5\ \ \Lambda\ 0.1,0.1,0.\ \ 9,0.9\right|}{0.0000001\ +\left|0.1,0.1,0.\ \ 9,0.9\right|}$$

$$= \frac{\left|0.1,0.1,0.\ \ 9,0.\ 5\right|}{2.0000001} = \frac{\left|1.6\right|}{2.0000001} = 0.7999$$

$$\mathrm{MAF(1)} = \frac{\left|0.1,0.5,0.\ \ 9,0.5\ \ \Lambda\ 0.1,0.1,0.\ \ 9,0.9\right|}{\left|0.1,0.5,0.\ \ 9,0.5\right|}$$

$$= \frac{\left|0.1,0.1,0.\ \ 9,0.5\right|}{2.0} = \frac{\left|1.6\right|}{2.0} = 0.8000$$

ACF (1) has same category as of category of nucleotide pair, MFI (1) is greater than rho, so

$$\mathrm{WFI(1)} = 1*(0.1,0.5,0.9,0.5)\ \Lambda\ (0.1,0.1,0.9,0.9)$$
$$+(1-1)(0.1,0.1,0.9,0.9) = 0.1,0.1,0.9,0.5$$

$$\mathrm{WFI(1)} = 0.1,0.1,0.9,0.5\ ,\ \mathrm{CFI(1)} = \mathrm{L}\ .$$

For nucleotide pair (TT)

Category to be trained: ILL (Illogical)

Fuzzy Equivalent: 0.2(T), 0.2(T)

Learning Input (LIN) = 0.2, 1-0.2, 1-0.2, 0.2

=0.2, 0.8, 0.8, 0.2 (Case2)

Rho (ρ) =0.5, $\mathrm{LIW(1)} = 0.1,0.1,0.8,0.5$

Table 2. Various cases for learning input normalization

	Condition	Learning Input	Category
Case 1	i ≠ n or i=n=0.1 and n<=0.5	$LIN_{i,n}$ = i, n, 1- i, 1-n e.g. $LIN_{0.1,0.1}$ = 0.1, 0.1, (1- 0.1), (1-0.1) $LIN_{0.1,0.1}$ = 0.1, 0.1, 0.9, 0.9 $LIN_{0.2,0.5}$ = 0.2, 0.5, (1- 0.2), (1-0.5) $LIN_{0.2,0.5}$ = 0.2, 0.5, 0.8, 0.5	Category=L (logical)
Case 2	i = n and 0.1> i, n <0.5	$LIN_{i,n}$ = i, 1-i, 1-n, n e.g. $LIN_{0.2,0.2}$= 0.2, (1-0.2), (1- 0.2), 0.2 $LIN_{0.2,0.2}$ = 0.2, 0.8, 0.8, 0.2 $LIN_{0.3,0.3}$= 0.3, (1-0.3), (1- 0.3), 0.3 $LIN_{0.3,0.3}$= 0.3, 0.7, 0.7, 0.3	Category=ILL (illogical)
Case 3	i=n=0.5	$LIN_{i,n}$ = i, i+0.1, n, n-0.1 e.g. $LIN_{0.5,0.5}$= 0.5, (0.5+1), 0.5, (0.5-0.1) $LIN_{0.5,0.5}$= 0.5, 0.6, 0.5, 0.4	Category=ILL (illogical)

$$ACF(1) = \frac{|0.2,0.8,0.8,0.2 \Lambda\ 0.1,0.1,0.8,0.5|}{0.0000001 + |0.1,0.1,0.8,0.5|}$$

$$= \frac{|0.1,0.1,0.8,0.2|}{1.5000001} = \frac{|1.2|}{1.5000001} = 0.9999 \quad ,$$

$$MAF(1) = \frac{|0.2,0.8,0.8,0.2 \Lambda\ 0.1,0.1,0.8,0.5|}{|0.2,0.8,0.8,0.2|}$$

$$= \frac{|0.1,0.1,0.8,0.2|}{2.0} = \frac{|1.2|}{2.0} = 0.6000$$

As category of nucleotide pair is the new category to be trained:

$$WFI(1) = 0.1,0.1,0.8,0.5 \ , \ WFI(2) = 0.2,0.8,0.8,0.2 .$$

$$CFI(1) = L \ , \ CFI(2) = ILL .$$

For nucleotide pair (CC)

Category to be trained: ILL (Illogical)

Fuzzy Equivalent: 0.4(C), 0.4(C)

Learning Input (LIN): 0.4, 1-0.4, 1-0.4, 0.4
=0.4, 0.6, 0.6, 0.4(Case2)

Rho (ρ) =0.5

$$LIW(1) = 0.1,0.1,0.6,0.5 \ , \ LIW(2) = 0.2,0.7,0.7,0.2$$

$$ACF(1) = \frac{|0.4,0.6,0.\ 6,0.4\ \Lambda\ 0.1,0.1,0.\ 6,0.5|}{0.0000001 + |0.1,0.1,0.\ 6,0.5|}$$

$$= \frac{|0.1,0.1,0.\ 6,0.4|}{1.3000001} = \frac{|1.2|}{1.3000001} = 0.9230$$

$$ACF(2) = \frac{|0.4,0.6,0.\ 6,0.4\ \Lambda\ 0.2,0.7,0.\ 7,0.2|}{0.0000001 + |0.2,0.7,0.\ 7,0.2|}$$

$$= \frac{|0.2,0.6,0.\ 6,0.2|}{1.8000001} = \frac{|1.6|}{1.8000001} = 0.8888$$

$$MAF(1) = \frac{|0.4,0.6,0.\ 6,0.4\ \Lambda\ 0.1,0.1,0.\ 6,0.5|}{|0.4,0.6,0.\ 6,0.4|}$$

$$= \frac{|0.1,0.1,0.\ 6,0.4|}{2.0} = \frac{|1.2|}{2.0} = 0.6000$$

$$MAF(2) = \frac{|0.4,0.6,0.\ \ 6,0.4\ \Lambda\ 0.2,0.7,0.\ \ 7,0.2|}{|0.4,0.6,0.\ \ 6,0.4|}$$

$$= \frac{|0.2,0.6,0.\ \ 6,0.2|}{2.0} = \frac{|1.6|}{2.0} = 0.8000$$

ACF (1) has the highest value and its category is Logical

which is not same as the category of input nucleotide pair, update the value of rho, that is, 0.600 (MAF (1)) +0.001(δ) =0.601, find the next highest value which is ACF (2) whose category is same as that of nucleotide pair, so update WIF(2):

$$\mathrm{WFI}(1) = 0.1,0.1,0.6,0.5 \ \mathrm{CFI}(1) = \mathrm{L}.$$

$$\mathrm{WFI}(2) = 1*(0.4,0.6,06,0.4) \Lambda (0.2,0.7,07,0.2)$$

$$+(1-1)(0.2,0.7,0.7,0.2) = 0.2,0.6,06,0.2,$$

$$\mathrm{CFI}(2) = \mathrm{ILL}.$$

For nucleotide pair (UA)

Category to be trained: L (Logical)

Fuzzy Equivalent: 0.5(U), 0.1(A)

Learning Input (LIN): 0.5, 0.1, 1-0.5, 1-0.1
=0.5, 0.1, 0.5, 0.9(Case1)

Rho (ρ) =0.601,

$$\mathrm{LIW}(1) = 0.1,0.1,0.6,0.5 \ , \mathrm{LIW}(2) = 0.2,0.6,0.6,0.2$$

$$\mathrm{ACF}(1) = \frac{\left|0.5,0.1,0.\ \ 5,0.9 \ \Lambda\ 0.1,0.1,0.\ \ 6,0.5\right|}{0.0000001 + \mid 0.1,0.1,0.\ \ 6,0.5 \mid}$$

$$= \frac{\left|0.1,0.1,0.\ \ 5,0.5\right|}{1.3000001} = \frac{\left|1.2\right|}{1.3000001} = 0.9230$$

$$\mathrm{ACF}(2) = \frac{\left|0.5,0.1,0.\ \ 5,0.9 \ \Lambda\ 0.2,0.6,0.\ \ 6,0.2\right|}{0.0000001 + \mid 0.2,0.6,0.\ \ 6,0.2 \mid}$$

$$= \frac{\left|0.2,0.1,0.\ \ 5,0.2\right|}{1.6000001} = \frac{\left|1.0\right|}{1.6000001} = 0.6249$$

$$\mathrm{MAF}(1) = \frac{\left|0.5,0.1,0.\ \ 5,0.9 \ \Lambda\ 0.1,0.1,0.\ \ 6,0.5\right|}{\mid 0.5,0.1,0.\ \ 5,0.9 \mid}$$

$$= \frac{\left|0.1,0.1,0.\ \ 5,0.5\right|}{2.0} = \frac{\left|1.2\right|}{2.0} = 0.6000$$

$$\mathrm{MAF}(2) = \frac{\left|0.5,0.1,0.\ 5,0.9 \ \Lambda\ 0.2,0.6,0.\ 6,0.2\right|}{\mid 0.4,0.6,0.\ 6,0.4 \mid}$$

$$= \frac{\left|0.2,0.1,0.\ 5,0.2\right|}{2.0} = \frac{\left|1.0\right|}{2.0} = 0.5000$$

ACF (1) has the highest value and MAF (1) is less than rho, find the next highest value which is ACF (2) ,its corresponding MAF (2) is also less than rho, so add new WFI (3), that is, WFI(3)=LIN:

$$\mathrm{WFI}(1) = 0.1,0.1,06,0.5, \mathrm{WFI}(2) = 0.2,0.6,06,0.2$$

$$\mathrm{WFI}(3) = 0.5,0.1,0.5,0.9 \ , \mathrm{CFI}(1) = \mathrm{L},$$

$$\mathrm{CFI}(2) = \mathrm{ILL} \ , \ \mathrm{CFI}(3) = \mathrm{L}.$$

For nucleotide pair (UU)

Category to be trained: ILL (Illogical)

Fuzzy Equivalent: 0.5(U), 0.5(U)

Learning Input (AI): 0.5, 0.5+0.1, 0.5, 0.5-0.1
=0.5, 0.6, 0.5, 0.4(Case3)

Rho (ρ) =0.601, $\mathrm{LIW}(1) = 0.1,0.1,0.6,0.5$,

$$\mathrm{LIW}(2) = 0.2,0.6,0.6,0.2 \ , \mathrm{LIW}(3) = 0.5,0.1,0.5,0.9$$

$$\mathrm{ACF}(1) = \frac{\left|0.5,0.6,0.\ \ 5,0.4 \ \Lambda\ 0.1,0.1,0.\ \ 6,0.5\right|}{0.0000001 + \mid 0.1,0.1,0.\ \ 6,0.5 \mid}$$

$$= \frac{\left|0.1,0.1,0.\ \ 5,0.4\right|}{1.3000001} = \frac{\left|1.1\right|}{1.3000001} = 0.8461$$

$$\mathrm{ACF}(1) = \frac{\left|0.5,0.6,0.\ \ 5,0.4 \ \Lambda\ 0.1,0.1,0.\ \ 6,0.5\right|}{0.0000001 + \mid 0.1,0.1,0.\ \ 6,0.5 \mid}$$

$$= \frac{\left|0.1,0.1,0.\ \ 5,0.4\right|}{1.3000001} = \frac{\left|1.1\right|}{1.3000001} = 0.8461$$

$$\mathrm{ACF}(2) = \frac{\left|0.5,0.6,0.\ \ 5,0.4 \ \Lambda\ 0.2,0.6,0.\ \ 6,0.2\right|}{0.0000001 + \mid 0.2,0.6,0.\ \ 6,0.2 \mid}$$

$$= \frac{\left|0.2,0.6,0.\ \ 5,0.2\right|}{1.6000001} = \frac{\left|1.5\right|}{1.6000001} = 0.9375$$

$$\mathrm{ACF}(3) = \frac{\left|0.5,0.6,0.\ 5,0.4 \ \Lambda\ 0.5,0.1,0.\ 5,0.9\right|}{0.0000001 + \mid 0.5,0.1,0.\ 5,0.9 \mid}$$

$$= \frac{\left|0.5,0.1,0.\ 5,0.4\right|}{2.0000001} = \frac{\left|1.5\right|}{2.0000001} = 0.7499$$

$$\mathrm{MAF}(1) = \frac{\left|0.5,0.6,0.\ \ 5,0.4 \ \Lambda\ 0.1,0.1,0.\ \ 6,0.5\right|}{\mid 0.5,0.6,0.\ \ 5,0.4 \mid}$$

$$= \frac{\left|0.1,0.1,0.\ \ 5,0.4\right|}{2.0} = \frac{\left|1.1\right|}{2.0} = 0.5500$$

Table 3. Various conditions for generating preprocessor output.

Preprocessor Input		Condition	Preprocessor Output
PPI(i, n)	1	i ≠ n or i=n=0.1 and n<=0.5	PPO(i, n) = i, n, 1- i, 1-n e.g. PPO(0.1, 0.1) = 0.1, 0.1, (1- 0.1), (1-0.1) PPO(0.1, 0.1) = 0.1, 0.1, 0.9, 0.9 PPO(0.2, 0.5) = 0.2, 0.5, (1- 0.2), (1-0.5) PPO(0.2, 0.5) = 0.2, 0.5, 0.8, 0.5
	2	i = n and 0.1> i, n <0.5	PPO(i, n) = i, 1-i, 1-n, n e.g. PPO(0.2, 0.2) = 0.2, (1-0.2), (1- 0.2), 0.2 PPO(0.2, 0.2) = 0.2, 0.8, 0.8, 0.2 PPO(0.3, 0.3) = 0.3, (1-0.3), (1- 0.3), 0.3 PPO(0.3, 0.3) = 0.3, 0.7, 0.7, 0.3
	3	i=n=0.5	PPO(i, n) = i, i+0.1, n, n-0.1 e.g. PPO(0.5, 0.5) = 0.5, (0.5+1), 0.5, (0.5- 0.1) PPO(0.5, 0.5) = 0.5, 0.6, 0.5, 0.4

$$\mathrm{MAF}(2) = \frac{|0.5,0.6,0.5,0.4 \Lambda\ 0.2,0.6,0.6,0.2|}{|\,0.5,0.6,0.5,0.4\,|}$$

$$= \frac{|0.2,0.6,0.5,0.2|}{2.0} = \frac{|1.5|}{2.0} = 0.7500$$

$$\mathrm{MAF}(3) = \frac{|0.5,0.6,0.5,0.4 \Lambda\ 0.5,0.1,0.5,0.9|}{|\,0.5,0.6,0.5,0.4\,|}$$

$$= \frac{|0.5,0.1,0.5,0.4|}{2.0} = \frac{|1.5|}{2.0} = 0.7500$$

ACF (2) has the highest value, whose category is Illogical which is same as the category of input nucleotide pair, update WFI (2):

$$\mathrm{WFI}(2) = (0.5,0.6,0.5,0.4)\ \Lambda\ (0.2,0.6,0.6,0.2)$$

$$= 0.2,0.6,0.5,0.2 \ \text{So WFI}(1) = (0.1,0.1,0.6,0.5)$$

$$\mathrm{WFI}(3) = 0.5,0.1,0.5,0.9,\ \mathrm{CFI}(1) = \mathrm{L},$$
$$\mathrm{CFI}(2) = \mathrm{ILL},\ \mathrm{CFI}(3) = \mathrm{L}.$$

$$\mathrm{WFI}(3) = 0.5,0.1,05,0.9.$$

The WFI, CFI generated for various learning input is shown in Table 1.Once the network has been trained; the inference of patterns, logical or illogical, that is, the categories to which the patterns belong may be easily computed. This is accomplished by subjecting DNA input to CIF function through pre-processor. Various conditions for generating pre-processed output are shown in Table 3.

Category inference function (CIF)

$$\mathrm{CIFj} = \frac{|\mathrm{PPO}\ \Lambda\ \mathrm{WFIj}|}{|\mathrm{WFIj}|} \qquad (7)$$

Computation involved in finding CIF for DNA Input (TG) is shown as:

If CIF (1)/ CIF (3) is greater than CIF (2) then greatest inferred category (GIF) is CIF (1) /CIF (3), so the category inferred is logical, else if CIF (2) is greater than CIF (1) and CIF (3) then greatest inferred category(IC) is illogical.

For DNA input TG

Pre-processor Input (PPI): 0.2(T), 0.3(G)

Pre-processor Output (PPO): 0.2, 0.3, 0.8, 0.7

$$\mathrm{CIF}(1) = \frac{|0.2,0.3,0.\quad 8,0.7\quad \Lambda\ \ 0.1,0.1,0.\quad 6,0.5\ |}{|\ 0.1,0.1,0.\quad 6,0.5\ \ |}$$

$$= \frac{|0.1,0.1,0.\quad 6,0.5\ |}{1.3} = \frac{|1.3\ |}{1.3} = 1.0000$$

$$CIF(2) = \frac{|0.2,0.3,0.8,0.7 \Lambda\ 0.2,0.6,0.5,0.2|}{|0.2,0.6,0.5,0.2|}$$

$$= \frac{|0.2,0.3,0.5,0.2|}{1.5} = \frac{|1.2|}{2.0} = 0.6000$$

$$CIF(3) = \frac{|0.2,0.3,0.8,0.7 \Lambda\ 0.5,0.1,0.5,0.9|}{|0.5,0.1,0.5,0.9|}$$

$$= \frac{|0.2,0.1,0.5,0.7|}{2.0} = \frac{|1.5|}{2.0} = 0.7500$$

CIF(1) has the highest value whose category is logical so the corresponding seven consecutive nucleotide base from TG in the DNA sample is chosen as single logical sequence i.e.0.2,0.3,0.2,0.3,0.2,0.3,0.1 and DNA inputs whose category is illogical, two consecutive similar nucleotide base is considered as an illogical sequence as shown in Table 4.

Logical sequence (LS):

$$LSp,s,k = Lseq\ p,s,1, Lseq\ p,s,2,..., Lseq\ p,s,k$$

$$\text{Where} \quad p,s = 1 \text{ to } \infty \quad \text{and} \quad k = 1 \text{ to } 7 \tag{8}$$

The sequence that are logical in their category alone are fed to the discriminator (D1) where unique identification number(0.182464) is computed using the equation 9 as shown in Table 6 and the standard deviation of logical sequence is calculated and plotted using MATLAB (Sivanandam, 2006) to represent unique repeated logical sequence pictorially as in Figure 2.

$$D1_{p,s} = \sum_{k=1}^{7} k(Lseq_{p,s,k})^{k} \tag{9}$$

$$p,s = 1 \text{ to } \infty$$

$$e.g. D1_{1,3} = 1(0.4)^{1} + 2(0.2)^{2} + 3(0.4)^{3} + 4(0.1)^{4}$$

$$+5(0.1)^{5} + 6(0.1)^{6} + 7(0.1)^{7} = 0.672457$$

Illogical sequence (IS):

$$IS_{p,s} = ILseq\ s, ILseq\ s,..., ILseq\ \infty \tag{10}$$

The sequence that are illogical in their category are fed to the discriminator (D2) where identification number is computed as shown in Table 5a and b using the equation:

$$D2_{p,s} = ILseq\ s^{m}$$

$$\text{where } p,s,m = 1 \text{ to } \infty$$

$$m = \text{Number of times nucleotide base is repeated}. \tag{11}$$

$$e.g. \quad D2_{1,2} = (0.2)^{3} = 0.008000$$

The discriminator outputs of both D1, D2 are used to identify the location of mutation in the given sample as thus discussed.

DNA SAMPLE: HUMAN-1 [BASE PAIR =32, SEQUENCE =25]

>AB000263 |acc=AB000263|descr=Homo sapiens mRNA for prepro cortistatin like peptide, complete cds.|len=368
AATGTGTTGTGTGACCCCTCAAAATCTCTCAAATGTG
TTTTTACACTCCGTTGGTAATATGGAATGTGTTAAAGT
TGCTACCCGGGGTTTTTTAATGTGTCTCT
TGTGACCCCTCAAAATCTCTCAAATGTGTTTTTACACT
CCGTTGGTAATATGGAATGTGTTAAAGTTGCTACCCG
GGGTTTTTTAATGTGTCTCT

IDENTIFCIATION OF MUTATION IN THE SAMPLE

Mutation (Charles, 2007) is a change of DNA sequence within a gene or chromosome of an organism resulting in the creation of a new character or trait not found in the parental type .The mutation (Mark and Marcus, 2007) results when a change occurs in a chromosome, either through an alteration in the nucleotide sequence of the DNA coding for a gene or through a change in the physical arrangement of a chromosome.

Mutations (Graham, 2007) that result in missing DNA are called deletions. These can be small, or longer deletions that affect a large number of genes on the chromosome. Deletions can also cause frame-shift mutations. Mutations (Richard et al., 1998 that result in the addition of extra DNA are called insertions. Insertions can also cause frame-shift mutations, and generally result in a non-functional protein. In an inversion mutation, an entire section of DNA is reversed. A small inversion may involve only a few bases within a gene, while longer inversions involve large regions of a chromosome containing several genes.

Various types of mutation identification in human-1 sample

Before mutation

LS1/RS	LS2	IS1	IS1	IS1
AATGTGT	TGTGTGA	C	C	C

Table 4. Identification of logical and illogical sequence using CIF.

		DNA inputs of *human-1*											
PPI (Preprocessor Input)		**A,A***	**T,G***	**C,C****	**C,C****	**C,C****	**C,T***	**T,C***	**A,A***	**T,T****	**T,T****	**T,T****	**T,A***
		0.1,0.1	**0.2,0.3**	**0.4,0.4**	**0.4,0.4**	**0.4,0.4**	**0.4,0.2**	**0.2,0.4**	**0.1,0.1**	**0.2,0.2**	**0.2,0.2**	**0.2,0.2**	**0.2,0.1**
PPO Preprocessor Output)		0.1,0.1, 0.9,0.9	0.2,0.3, 0.8,0.7	0.4,0.6, 0.6,0.4	0.4,0.6, 0.6,0.4	0.4,0.6, 0.6,0.4	0.4,0.2, 0.6,0.8	0.2,0.4, 0.8,0.6	0.1,0.1, 0.9,0.9	0.2,0.8, 0.8,0.2	0.2,0.8, 0.8,0.2	0.2,0.8, 0.8,0.2	0.2,0.1, 0.8,0.9
WFI	WFI(1) / CFI(1)-L	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5	0.1,0.1, 0.6,0.5
	WFI(2) / CFI(2)-ILL	0.2,0.6, 0.5,0.2	0.2,0.6, 0.5,0.2	0.2,0.6, 0.5,0.2	0.2,0.6, 0.5,0.2	0.2,0.6, 0.5,0.2	0.2,0.6, 0.5,0.2	0.2,0.6, 0.5,0.2	0.2,0.6, 0.5,0.2	0.2,0.6, 0.5,0.2	0.2,0.6, 0.5,0.2	0.2,0.6, 0.5,0.2	0.2,0.6, 0.5,0.2
	WFI(3) / CFI(3)-L	0.5,0.1, 0.5,0.9	0.5,0.1, 0.5,0.9	0.5,0.1, 0.5,0.9	0.5,0.1, 0.5,0.9	0.5,0.1, 0.5,0.9	0.5,0.1, 0.5,0.9	0.5,0.1, 0.5,0.9	0.5,0.1, 0.5,0.9	0.5,0.1, 0.5,0.9	0.5,0.1, 0.5,0.9	0.5,0.1, 0.5,0.9	0.5,0.1, 0.5,0.9
CIF	CIF(1)	1.0000	1.0000	0.9230	0.9230	0.9230	1.0000	1.0000	1.0000	0.7692	0.7692	0.7692	1.0000
	CIF(2)	0.6000	0.6000	1.0000	1.0000	1.0000	0.7333	0.8666	0.6000	1.0000	1.0000	1.0000	0.6666
	CIF(3)	0.8000	0.7500	0.7000	0.7000	0.7000	0.9000	0.7000	0.8000	0.5000	0.5000	0.5000	0.8500
GIC		CIF(1)	CIF(1)	CIF(2)	CIF(2)	CIF(2)	CIF(1)	CIF(1)	CIF(1)	CIF(2)	CIF(2)	CIF(2)	CIF(1)
IC	LOGICAL	L	L				L	L	L				L
	ILLOGICAL			ILL	ILL	ILL				ILL	ILL	ILL	
Categorized Sequence		0.1,0.1,0.2,0.3,0.2, 0.3,0.2	0.2,0.3,0.2,0.3,0.2,0.3, 0.1	0.4,0.4	0.4,0.4	0.4,0.4	0.4,0.2,0.4,0.1,0.1,0.1, 0.1	0.2,0.4,0.2, 0.4,0.2,0.4, 0.1	0.1,0.1,0.2, 0.3,0.2,0.3, 0.2	0.2,0.2	0.2,0.2	0.2,0.2	0.2,0.1,0.4, 0.1,0.4,0.2, 0.4

CIF, Category inference function; ***,** condtion 1; ****,** condition 2, *****,** condition 3; **IC,** inferred category.

LS3 LS4 LS5/RS IS2 IS2
CTCAAA TCTCTCA AATGTGT T T
IS2 LS6 LS7 LS8
T TACACTC CGTTGGT AATATGG
LS9/RS LS10 LS11 IS3 IS3
AATGTGT TAAAGTT GCTACCC G G
IS3 LS12 LS13/RS LS14
G GTTTTTT AATGTGT CTCTXXX

Case 1

After point mutation in the sample

LS1/RS LS2 IS1 IS1 IS1
AATGTGT TGTGTGA C C C
LS3 LS4 LS5/RS IS2 IS2
CTCA **C** AA TCTCTCA AATGTG T T
IS2 LS6 LS7 LS8
T TACACTC CGTTGGT AATATGG
LS9/RS LS10 LS11 IS3
AATGTGT TAAAGTT GCTACCC G

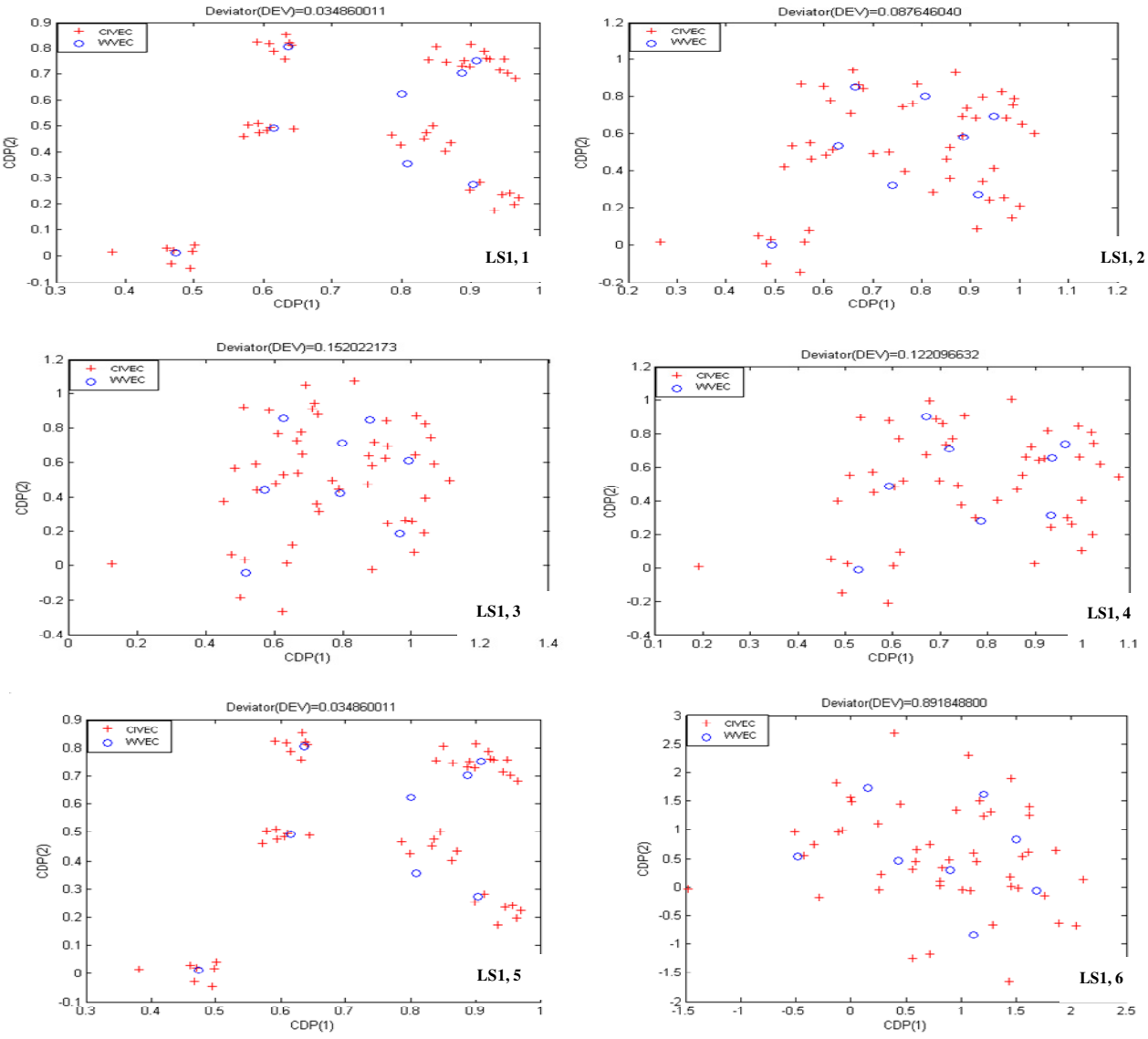

Figure 2. MATLAB output for logical sequence ($LS_{(1,1)}$-$LS_{(1,6)}$) showing $LS_{(1,1)}$ and $LS_{(1,5)}$ are unique. CDP, Clustered data points; CIVEC, cluster of input vectors; WVEC, weight vectors.

IS3 IS3 LS12 LS13/RS LS14
G G GTTTTTT AATGTGT CTCTXXX
In case 1 the point mutation occurred in logical sequence ($LS_{1,3}$) by the mutant C that can be identified with the change in identification number of $LS_{1,3}$ where identification number of illogical sequence remains unaltered as in Table 6.

Result

Change in polypeptide sequence might change the shape or function of the protein, depending on where in the sequence occurs.

Case 2

After frame shift mutation [Insertion] in the sample:

LS1/RS LS2 IS1 IS1 IS1
AATGTGT TGTGTGA C C C
LS3 LS4 LS5/RS IS2 IS2
CTCAAAA TCTCTCA AATGTGT T T
IS2 LS6 LS7 LS8
T TACACTC CGTTGGT AATATGG

Table 5a. Discriminator (d2) outputs for categorized IS from non-mutated human 1 sample.

Illogical Sequence ($IS_{p,s}$)	Number of time sequence repeated(m)	Human (p)	Sequence (s)	$ILseq_s$	Identification number ($D2_{p,s}$)
$IS_{1,1}$	1	1	1	0.4	0.064000
$IS_{1,1}$	2	1	1	0.4	
$IS_{1,1}$	3	1	1	0.4	
$IS_{1,2}$	1	1	2	0.2	0.008000
$IS_{1,2}$	2	1	2	0.2	
$IS_{1,2}$	3	1	2	0.2	
$IS_{1,3}$	1	1	3	0.3	0.027000
$IS_{1,3}$	2	1	3	0.3	
$IS_{1,3}$	3	1	3	0.3	

Table 5b. Discriminator (d1) outputs for categorized logical sequence from non-mutated human1 sample.

Logical sequence ($LS_{p,s}$)	Human (p)	Sequence (s)	$LSeq_{p,s,k}$							Identification number ($D1_{p,s}$)	
			k=1	k=2	k=3	k=4	k=5	k=6	k=7		
$LS_{1,1}$	1	1	0.1	0.1	0.2	0.3	0.2	0.3	0.2	**0.182464**	0.182464 (Unique identification number)
$LS_{1,2}$	1	2	0.2	0.3	0.2	0.3	0.2	0.3	0.1	0.442375	
$LS_{1,3}$	1	3	0.4	0.2	0.4	0.1	0.1	0.1	0.1	0.672457	
$LS_{1,4}$	1	4	0.2	0.4	0.2	0.4	0.2	0.4	0.1	0.672577	
$LS_{1,5}$	1	5	0.1	0.1	0.2	0.3	0.2	0.3	0.2	**0.182464**	
$LS_{1,6}$	1	6	0.2	0.1	0.4	0.1	0.4	0.2	0.4	0.475453	
$LS_{1,7}$	1	7	0.4	0.3	0.2	0.2	0.3	0.3	0.2	0.627014	
$LS_{1,8}$	1	8	0.1	0.1	0.2	0.1	0.2	0.3	0.3	0.151905	
$LS_{1,9}$	1	9	0.1	0.1	0.2	0.3	0.2	0.3	0.2	**0.182464**	
$LS_{1,10}$	1	10	0.2	0.1	0.1	0.1	0.3	0.2	0.2	0.236024	
$LS_{1,11}$	1	11	0.3	0.4	0.2	0.1	0.4	0.4	0.4	0.731645	
$LS_{1,12}$	1	12	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.412474	
$LS_{1,13}$	1	13	0.1	0.1	0.2	0.3	0.2	0.3	0.2	**0.182464**	

LS9/RS **LS10** **LS11** **LS12**
AATGTGT TAAAGTT GCTACCC G **C** GGGTT
IS3 **IS3** **IS3** **LS13** **LS14**
T T T TAATGTG TCTCTXX

In case 2 the frame shift mutation (insertion) occurred in one of the $IS_{1,3}$ by the mutant C which alters both the logical sequence ($LS_{1,12}$) and illogical sequence ($IS_{1,3}$) that can be identified by the change in identification number of both logical sequence ($LS_{1,12}$) and illogical sequence ($IS_{1,3}$) as in Table 7.

Result

Change in polypeptide sequence might change the shape or function of the protein, depending on where in the sequence occurs.

Case 3

After point mutation [neutral or silent] in the sample

LS1/RS **LS2** **IS1** **IS1** **IS1**
AATGTGT TGTGTGA C C C
LS3 **LS4** **LS5/RS** **IS2** **IS2**
CTCAAAA TCTCTCA AATGTGT T T
IS2 **LS6** **LS7** **LS8**
T TACACTC CGTTGGT AATATGG
LS9/RS **LS10** **LS11** **IS3**
AATGTGT TAAAGTT GCTACCC G

Table 6. Point mutation.

Logical Sequence (LS)	Identification number (before mutation)	Identificatio number (after mutation)
$LS_{1,1}$	0.182464	0.182464
$LS_{1,2}$	0.442375	0.442375
$LS_{1,3}$	0.672457	**0.723607**
$LS_{1,4}$	0.672577	0.672577
$LS_{1,5}$	0.182464	0.182464
$LS_{1,6}$	0.475453	0.475453
$LS_{1,7}$	0.627014	0.627014
$LS_{1,8}$	0.151905	0.151905
$LS_{1,9}$	0.182464	0.182464
$LS_{1,10}$	0.236024	0.236024
$LS_{1,11}$	0.731645	0.731645
$LS_{1,12}$	0.412474	0.412474
$LS_{1,13}$	0.182464	0.182464
Illogical sequence (IS)	**Identification number (before mutation)**	**Identification number (after mutation)**
$IS_{1,1}$	0.064000	0.064000
$IS_{1,2}$	0.008000	0.008000
$IS_{1,3}$	0.027000	0.027000

Table 7. Frame shift mutation (insertion).

Logical Sequence (LS)	Identification number (before mutation)	Identification number (after mutation)
$LS_{1,1}$	0.182464	0.182464
$LS_{1,2}$	0.442375	0.442375
$LS_{1,3}$	0.672457	0.672457
$LS_{1,4}$	0.672577	0.672577
$LS_{1,5}$	0.182464	0.182464
$LS_{1,6}$	0.475453	0.475453
$LS_{1,7}$	0.627014	0.627014
$LS_{1,8}$	0.151905	0.151905
$LS_{1,9}$	0.182464	0.182464
$LS_{1,10}$	0.236024	0.236024
$LS_{1,11}$	0.731645	0.731645
$LS_{1,12}$	0.412474	0.7459736
$LS_{1,13}$	0.182464	0.243464
Illogical sequence (IS)	**Identification number (before mutation)**	**Identification number (after mutation)**
$IS_{1,1}$	0.064000	0.064000
$IS_{1,2}$	0.008000	0.008000
$IS_{1,3}$	0.027000	0.008000

Table 8. point mutation neutral or silent.

Logical sequence (LS)	Identification number (before mutation)	Identification number (after mutation)
$LS_{1,1}$	0.182464	0.182464
$LS_{1,2}$	0.442375	0.442375
$LS_{1,3}$	0.672457	0.672457
$LS_{1,4}$	0.672577	0.672577
$LS_{1,5}$	0.182464	0.182464
$LS_{1,6}$	0.475453	0.475453
$LS_{1,7}$	0.627014	0.627014
$LS_{1,8}$	0.151905	0.151905
$LS_{1,9}$	0.182464	0.182464
$LS_{1,10}$	0.236024	0.236024
$LS_{1,11}$	0.731645	0.731645
$LS_{1,12}$	0.412474	0.412474
$LS_{1,13}$	0.182464	0.182464
Illogical sequence (IS)	**Identification number (before mutation)**	**Identification number (after mutation)**
$IS_{1,1}$	0.064000	0.064000
$IS_{1,2}$	0.008000	0.008000
$IS_{1,3}$	0.027000	**0.008100**

IS3 IS3 IS3 LS12 LS13/RS
G G G GTTTTTT AATGTGT
LS14
CTCTXXX

Result

No change in polypeptide sequence, possible consequence for the organism =none**.** In case 3, the point mutation is occurred in same $IS_{1,3}$ as case 2 but with mutant G that only alters the illogical sequence ($IS_{1,3}$) and not any of the LS that can be identified only using the change in identification number of illogical sequence ($IS_{1,3}$) Table 8.

Case 4

After Frame shift mutation in the sample

LS1/RS LS2 IS1 IS1 IS1
AATGTGT TGTGTGA C C C
LS3 LS4 LS5/RS IS2 IS2
CTCAAAA TCTCTCA AATGTGT T T
IS2 LS6 LS7 LS8
T TACACTC CGTTGGT AATATGG
LS9/RS LS10 LS11 IS3 IS3
AATGTGT TAAAGTT GCTACCC G G
IS3 LS12 LS13/RS LS14
GGTTT TTA ATGTGTC TCTXXX

In case 4 the frame mutation [deletion] occurred in logical sequence ($LS_{1,\ 12)}$ by the removal of mutant T and can be identified with the change in identification number of logical sequence ($LS_{1,\ 12)}$ with no alteration in any of the illogical sequence as in Table 9 .

Result

Change in polypeptide sequence might change the shape or function of the protein, depending on where in the sequence occurs.

Case 5

In case 5, the inversion mutation occurred in logical sequence ($LS_{1,\ 10}$) by replacing TAAAGTT with mutant TTGAAAT that can be identified with the change in identification number of logical sequence ($LS_{1,\ 10}$) alone with no alteration in any of the IS as in Table 10.

After inversion mutation in the sample

LS1/RS LS2 IS1 IS1 IS1

Table 9. Frame shift mutation deletion.

Logical sequence (LS)	Identification number (before mutation)	Identification number (after mutation)
$LS_{1,1}$	0.182464	0.182464
$LS_{1,2}$	0.442375	0.442375
$LS_{1,3}$	0.672457	0.672457
$LS_{1,4}$	0.672577	0.672577
$LS_{1,5}$	0.182464	0.182464
$LS_{1,6}$	0.475453	0.475453
$LS_{1,7}$	0.627014	0.627014
$LS_{1,8}$	0.151905	0.151905
$LS_{1,9}$	0.182464	0.182464
$LS_{1,10}$	0.236024	0.236024
$LS_{1,11}$	0.731645	0.731645
$LS_{1,12}$	0.412474	**0.410945**
$LS_{1,13}$	0.182464	**0.291402**
Illogical sequence (IS)	**Identification number (before mutation)**	**Identification number (after mutation)**
$IS_{1,1}$	0.064000	0.064000
$IS_{1,2}$	0.008000	0.008000
$IS_{1,3}$	0.027000	0.027000

Table 10. Inversion mutation.

Logical sequence (LS)	Identification number (before mutation)	Identification number (after mutation)
$LS_{1,1}$	0.182464	0.182464
$LS_{1,2}$	0.442375	0.442375
$LS_{1,3}$	0.672457	0.672457
$LS_{1,4}$	0.672577	0.672577
$LS_{1,5}$	0.182464	0.182464
$LS_{1,6}$	0.475453	0.475453
$LS_{1,7}$	0.627014	0.627014
$LS_{1,8}$	0.151905	0.151905
$LS_{1,9}$	0.182464	0.182464
$LS_{1,10}$	0.236024	0.361546
$LS_{1,11}$	0.731645	0.731645
$LS_{1,12}$	0.412474	0.412474
$LS_{1,13}$	0.182464	0.182464
Ilogical sequence (IS)	**Identification number (before mutation)**	**Identification number (after mutation)**
$IS_{1,1}$	0.064000	0.064000
$IS_{1,2}$	0.008000	0.008000
$IS_{1,3}$	0.027000	0.027000

AATGTGT TGTGTGA C C C
LS3 LS4 LS5/RS IS2 IS2
CTCAAAA TCTCTCA AATGTGT T T
IS2 LS6 LS7 LS8
T TACACTC CGTTGGT AATATGG
LS9/RS LS10 LS11 IS3
AATGTGT **TTGAAAT** GCTACCC G
IS3 IS3 LS12 LS13/RS LS14
G G GTTTTTT AATGTGT CTCTXXX

CONCLUSION

As an attempt to automate the genetic finger printing the Neural-fuzzy pattern recognition system (NFPR) discussed in the above work assists forensic scientists by generating unique identification number for individuals from their DNA sample. The proposed system also helps to identify the location of occurrence mutation in the given mutated DNA sample, for instance, gene mutations which triggers hereditary nonpolyposis colorectal cancer (HNPCC) tumor that could not be detected even by PCR-SSCP can be easily detected by subjecting the sample to gene sequencing process and analyzed using above system.

Further development can be extended by training patterns in DNA protein that can be represented by suitable fuzzy equivalent in order to classify and predict the protein structure in the protein folding problem. The above technique can be used in the areas where feature extraction is to be done in genetic engineering with suitable modification.

ACKNOWLEDGEMENTS

We like to thank Senior Scientist, Dr. K. Thangarasu of Bio-Rad Laborataries, Banaglore, India and Dr. K. Somasundaram of Amirtha University, Coimbatore, India for their assistance in conducting the experiment.

Abbreviations: DNA, DEOXYRIBONUCLEIC acid; **RNA,** ribonucleic acid; **NFPR,** neural-fuzzy pattern recognition; **NFRM,** neural- fuzzy resonance mapping; **A,** adenine; **T,** thymine; **G,** guanine; **C,** cytosine; **U,** uracil; **ACF,** activation function; **MAF,** match function; **CFI,** category for Inference; **CIF,** category inference function; I**S,** illogical sequence; **CDP,** clustered data points; **CIVEC,** cluster of input vectors; **WVEC,** weight vectors; **HNPCC,** hereditary nonpolyposis colorectal cancer; **SL,** logical sequence.

REFERENCES

Mark D.Evans and Marcus S.Cooke (2007). Oxidative Damage to Nucleic Acids, Springer science press, New York.

Advances in Neural Networks issn (2006). Third international symposium on neural networks, Springer Berlin Heidelberg, New York publications.

Andreas DB (2001). Bioinformatics-A practical Guide to the Analysis of genes and proteins, second edition, A John wiley & sons, Inc., Publication.

Ayala-Gross JA (2001). DNA Analysis: The best method for Human Identifications, National University, San Diego.

Carpenter GA, Grossberg (1987). A Massively Parallel Architecture for a self-organizing Neural Pattern Recognition Machine, Computer Vision, Graphics and Image Processing, 37: 54-115.

Carpenter GA, Grossberg S, Reynolds JH (2010). "ARTMAP: Supervised Real Time Learning and Classification of Non-stationary Data by a Self- organizing Neural Network", 4: 565-588.

Charles LV (2007). New developments in Mutation Research, Nova science publishers Inc New York.

David EN (2008). DNA Evidence and Forensic science- 2008 facts on file, Inc. http://www.factsonfile.com.

David WM (2004). Bioinformatics Sequence and Genome analysis-Second Edition, Cold Spring Harbor Laboratory Press, New York.

Des H, Willie T (2000). Bioinformatics Sequence, Structure and data banks, Oxford University Press.

Donald RT (2005). The Pattern Recognition Basis of Artificial Intelligence", IEEE Press, New York. Computational Intelligence and Bio inspired Systems, 8th international work conference on artificial neural networks, iwann-2005proceedings. Elsevier Inc., p. 117.

Graham RT (2007). Laboratory methods for the detection of mutations and polymorphisms in DNA, CRC Press, 2007 - Science.

Joe N, John FF (1999). Crime Science Methods of Forensic Detection, 1999. University Press of Kentucky.

John H, Anders K, Richard GP (2008). Introduction to the Theory of Neural Computation, Addison Wesley, Redwood City, A.

John OS, Brent ET (2005). Rape Investigation Hand book, 2005,

Michael RB (2007). Bioinformatics for geneticists, Second Edition, John Wiley & Sons Ltd.

Norah R, Keith I (2002). An Introduction forensic DNA Analysis, 2002-CRC Press.

Phipps A, Lawrence J, Geert-De-Soete (1996).. Clustering and Classification, World Scientific, River Edge, NJ.

Richard GHC, Edward E, Sue F (1998). Mutation detection, IRL Press at Oxford University Press.

Richard OD, Peter EH, David GS (2006). Pattern classification-Second Edition, John Wiley and sons.

Robert S (2007). Pattern Recognition: Statistical, Structural and Neural Approaches, 2007, John Wiley and sons.

Sivanandam SN (2006). Introduction to neural networks and MATLAB-6.0", Tata McGraw-Hill publishing company.

Stephen JH, Jack DC, Giles CL (2009). Advances in Neural Information Processing Systems, volume 5, Morgan Kaufmann San Mateo CA.

Stephen K, David DW (2003). Introduction to Bioinformatics A Theoretical and Practical Approach, Human Press Inc.