A NEURAL NETWORK APPROACH FOR THE PRECISE PATTERN RECOGNITION OF HUMAN DNA

B. Mukunthan*

Faculty, Department of Master of Computer Applications, SVS Institute of Computer Applications, SVS College of Engineering, Ann University, Coimbatore-109, Tamil Nadu, India, E-mail: mukunth_bmk@yahoo.co.in

Abstract: The primary goal of bio informatics and neural networks solely is to increase our understanding of biological processes and focus on developing and applying computationally intensive techniques (e.g., pattern recognition, data mining, machine learning algorithms, and visualization) to achieve this goal. The neural networks exhibit characteristics such as mapping capabilities or pattern association, generalization, robustness, fault tolerance, parallel and high speed information processing. Neural networks learn by examples they can therefore be trained with known examples of a problem to ‘acquire’ knowledge about it. Once appropriately trained, the network can be put to effective use in solving ‘unknown’ or ‘untrained’ instances of the problem. The perfect blend made of bioinformatics and neural networks results in efficient pattern analysis techniques. The conventional techniques and algorithms employed by forensic scientists to assist in the identification of individuals on the basis of their respective DNA profiles involves more computational steps and mathematical formulas that leads to more time and space complexity resulting in complicated and less efficient algorithms which can be short cut by emerging Artificial Neural Network approach.

Keywords: DNA Sequence Format, Competitive Learning, Simplified Fuzzy ARTMAP, Input Generator, Preprocessor, Separator, and Discriminator.

1. INTRODUCTION

Neural networks [1], which are simplified models of the biological neuron system, is a massively parallel distributed processing 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 Network architectures [2] have been classified into various types based on their learning mechanisms and other features. Some classes of Neural Network refer to this learning process as training and the ability to solve a problem using the knowledge acquired as inference.

Neural Networks exhibit mapping capabilities, i.e., they can map input patterns to their associated output patterns. Neural Networks learn by examples. Thus, Neural Network architectures can be ‘trained’ with known examples of a problem before they are tested for their ‘inference’ capability on unknown instances of the problem. They can, therefore, identify new objects previously untrained. Neural Networks possess the capability to generalize. They can predict new outcomes from past trends. Neural Networks are robust systems and are fault tolerant. They can therefore, recall full patterns from incomplete, partial or noisy patterns.

Neural Networks can process information in parallel, at high speed, and in a distributed manner. Learning methods in neural networks are classified as Supervised Learning, Unsupervised Learning, Reinforce Learning, Hebbian Learning, Gradient descent Learning, Competitive Learning [3], Stochastic Learning. 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 [4].

Adaptive resonance theory employs a new principle of self organization based on Competitive learning. Adaptive resonance theory nets are designed to be both stable and plastic. Neural networks suitable particularly for pattern classification problems in realistic environment is Simplified Fuzzy ARTMAP [5] it is a vast simplification of fuzzy ARTMAP which has reduced computational overhead and architectural redundancy when compared to its predecessor.

Bio informatics [6] is the application of information technology and computer science to the field of molecular biology. Its primary use has been in genomics and genetics, particularly in those areas of genomics involving large-scale DNA sequencing. Deoxyribonucleic acid (DNA) [7] is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms and some viruses.
The main role of DNA molecules is the long-term storage of information. DNA is often compared to a set of blueprints or a recipe, or a code, since it contains the instructions needed to construct other components of cells, such as proteins and RNA molecules.

2. DNA PROFILING

DNA profiling (also called DNA testing, DNA typing, or genetic fingerprinting) is a technique employed by forensic scientists to assist in the identification of individuals on the basis of their respective DNA profiles. DNA profiles are encrypted sets of numbers that reflect a person’s DNA makeup, which can also be used as the person’s identifier. DNA profiling should not be confused with full genome sequencing it is used in, for example, parental testing and rape investigation. Forensic science (often shortened to forensics) is the application of a broad spectrum of sciences to answer questions of interest to a legal system. This may be in relation to a crime or a civil action. Although 99.9% of human DNA sequences are the same in every person, enough of the DNA is different to distinguish one individual from another. DNA profiling uses repetitive (“repeat”) sequences that are highly variable, called variable number tandem repeats (VNTR). VNTRs loci are very similar between closely related humans, but so variable that unrelated individuals are extremely unlikely to have the same VNTRs.

3. DNA SEQUENCING

Knowledge of DNA sequences has become indispensable for basic biological research. 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. The sequence information is analyzed to determine genes that encode polypeptides, RNA genes, regulatory sequences, structural motifs, and repetitive sequences. 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. The rapid speed of sequencing attained with modern DNA sequencing technology has been instrumental in the sequencing of the human genome, in the Human Genome Project. The practical aspects revolve around designing and optimizing sequencing projects, predicting project performance, troubleshooting experimental results, characterizing factors such as sequence bias and the effects of software processing algorithms, and comparing various sequencing methods to one another.

In this sense, it could be considered a branch of systems engineering or operations research. DNA sequencing theory addresses physical processes related to sequencing DNA and should not be confused with theories of analyzing resultant DNA sequences, e.g. sequence alignment. The term DNA sequencing refers to sequencing methods for determining the order of the nucleotide bases—adenine, guanine, cytosine, thymine and uracil in a molecule of DNA.

Single nucleotide poly-orphisms is a DNA sequence variation occurring when a single nucleotide — A, T, C, or G — in the genome (or other shared sequence) differs between members of a species (or between paired chromosomes in an individual). The genome is the entirety...
of an organism’s hereditary information which is encoded either in DNA or, for many types of virus, in RNA. For example, two sequenced DNA fragments from different individuals, AAGCCTA to AAGCTTA, contain a difference in a single nucleotide. When testing for father-motherhood scientists typically only need to look at 10 or 20 genomic regions like Single nucleotide polymorphism to determine relationship or lack thereof i.e. a tiny fraction of the human genome, which consists of three billion or so nucleotide.

3.1. Deoxyribonucleic Acid Sequence Formats

Various DNA Sequence Formats [12] available are: (1) Plain sequence format (2) EMBL format (3) GCG format (4) GCG-RSF (rich sequence format (5) GenBank format 4) IG format FASTA format

3.1.1 FASTA format

A sequence file in FASTA format can contain several sequences. Each sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line must begin with a greater-than (“>”) symbol in the first column. It is a text-based format for representing either nucleotide sequences or peptide sequences, in which base pairs or amino acids are represented using single-letter codes. The format also allows for sequence names and comments to precede the sequences. The FASTA format may be used to represent either single sequences or many sequences in a single file. A series of single sequences, concatenated, constitute a multi sequence file. It is common to end the sequence with an “*” (asterisk) character and leave a blank line between the description and the sequence.

3.1.2 Example sequence in FASTA format

>AB000263 |acc=AB000263|descr=Homo sapiens mRNA for propro cortistatin like peptide, complete cds.|len=368

ACAAGATGCCATTGTCCCCGGCC
TCCTGCTGCTGCTTCCCGGGG
CCACGCGCACCCTGGGCTGCC
CCTGAAGGTTGGGCCCCACCCGGCC
GAGACAGCGGACATATGCAGGAA
CGCGCGAAATAAGGAAACAGCAGC
CTCCTGACCTTTCTCTGCTGTGGGTT
TTGAGTGGACCCTCCGAGGCGATG
CCGCCGCCCTCATAGGAGAGG
AAGCTCGGGAGGGCGCCAGCGCG
CAGGAAGGGCGACCCCCACAGCA
ATCCGCCGCCCGGACAGAATGCC
CTGCAAGGAACTTTCTCTGGAAGAC
CTTCTCTCTCTGCTAAATAACCT
CACCCATGAATGCTACGGAAG
TTTAATACAGACCTGAA

4. BIO-NEURAL PATTERN RECOGNITION TOOL

Figure 4: Block Diagram of Bio-Neural Pattern Recognition Tool

4.1.1 Training Input Generation (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.

Training Inputs (TIN)

\[ TIN_{i,n} = I_1, I_2, ... , I_p \]  \hspace{1cm} (1)

where \( 0.1 \leq i \leq 0.5, 0.1 \leq n \leq 0.5 \) and \( p = 4 \)

Case 1: \( i \neq n \) or \( I = n = 0.1 \) Then

\[ TIN_{i,n} = i, n, I-n \]

Category-V (valid)

\[ TIN_{0.1,0.1} = 0.1, 0.1, (1-0.1), (1-0.1) \]
\[ TIN_{0.1,0.1} = 0.1, 0.1, 0.9, 0.9 \]
\[ TIN_{0.2,0.5} = 0.2, 0.5, (1-0.2), (1-0.5) \]
\[ TIN_{0.2,0.5} = 0.2, 0.5, 0.8, 0.5 \]
\[ TIN_{0.4,0.5} = 0.4, 0.5, (1-0.4), (1-0.5) \]
\[ TIN_{0.4,0.5} = 0.4, 0.5, 0.6, 0.5 \]
\[ TIN_{0.5,0.1} = 0.5, 0.1, (1-0.5), (1-0.1) \]
\[ TIN_{0.5,0.1} = 0.5, 0.1, 0.5, 0.9 \]
Case 2: \( i = n \) and \( i, n > 0.1 \) Then

\[
TIN_{i, n} = i, 1-i, 1-n, n
\]

Category-INV (invalid)

\[
TIN_{0.2, 0.2} = 0.2, (1-0.2), (1- 0.2), 0.2
\]

\[
TIN_{0.2, 0.2} = 0.2, 0.8, 0.8, 0.2
\]

\[
TIN_{0.3, 0.3} = 0.3, (1-0.3), (1- 0.3), 0.3
\]

\[
TIN_{0.3, 0.3} = 0.3, 0.7, 0.7, 0.3
\]

\[
TIN_{0.4, 0.4} = 0.4, (1-0.6), (1- 0.6), 0.4
\]

\[
TIN_{0.4, 0.4} = 0.4, 0.6, 0.6, 0.4
\]

\[
TIN_{0.5, 0.5} = 0.5, (1-0.5), (1- 0.5), 0.5
\]

\[
TIN_{0.5, 0.5} = 0.5, 0.5, 0.5, 0.5
\]

4.1.2. Ignition Function and Tracking Function

When coded input patterns from input generator are presented to NF-Processor all output nodes become active to varying degrees. The output ignition denoted by \( IGF_j \) referred to as the ignition function for the \( j^{th} \) output node. Where \( TIN \) is the training input and \( TIW_j \) is the corresponding training input weights.

(Tracking Function)

\[
TRF_j = \frac{TIN \cdot TIW_j}{|TIN|} \quad (4)
\]

The tracking function in association with the monitoring parameter decides on whether a particular output node is 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 [13], it is essential that it not only encodes the given input pattern but should also represent the same category as that of the input patterns.

The network is said to be in state of mismatch reset if the monitoring 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 monitoring parameter.

Table 1

<table>
<thead>
<tr>
<th>S.NO</th>
<th>TRAINING INPUTS</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( T_{50,5} \cdot 0.1 \cdot 0.9 \cdot 10 \cdot 0.9 )</td>
<td>Valid</td>
</tr>
<tr>
<td>2</td>
<td>( T_{50,5} \cdot 0.1 \cdot 0.5 \cdot 0.9 \cdot 0.5 )</td>
<td>Valid</td>
</tr>
<tr>
<td>3</td>
<td>( T_{50,5} \cdot 0.2 \cdot 0.1 \cdot 0.8 \cdot 0.9 )</td>
<td>Valid</td>
</tr>
<tr>
<td>4</td>
<td>( T_{50,5} \cdot 0.2 \cdot 0.0 \cdot 0.8 \cdot 0.2 )</td>
<td>Invalid</td>
</tr>
<tr>
<td>5</td>
<td>( T_{50,5} \cdot 0.2 \cdot 0.5 \cdot 0.8 \cdot 0.5 )</td>
<td>Valid</td>
</tr>
<tr>
<td>6</td>
<td>( T_{50,5} \cdot 0.5 \cdot 0.1 \cdot 0.7 \cdot 0.9 )</td>
<td>Valid</td>
</tr>
<tr>
<td>7</td>
<td>( T_{50,5} \cdot 0.5 \cdot 0.5 \cdot 0.7 \cdot 0.3 )</td>
<td>Invalid</td>
</tr>
<tr>
<td>8</td>
<td>( T_{50,5} \cdot 0.5 \cdot 0.5 \cdot 0.7 \cdot 0.5 )</td>
<td>Valid</td>
</tr>
<tr>
<td>9</td>
<td>( T_{60,5} \cdot 0.4 \cdot 0.1 \cdot 0.6 \cdot 0.9 )</td>
<td>Valid</td>
</tr>
<tr>
<td>10</td>
<td>( T_{60,5} \cdot 0.4 \cdot 0.6 \cdot 0.6 \cdot 0.4 )</td>
<td>Invalid</td>
</tr>
<tr>
<td>11</td>
<td>( T_{60,5} \cdot 0.4 \cdot 0.5 \cdot 0.6 \cdot 0.5 )</td>
<td>Valid</td>
</tr>
<tr>
<td>12</td>
<td>( T_{60,5} \cdot 0.5 \cdot 0.1 \cdot 0.5 \cdot 0.9 )</td>
<td>Valid</td>
</tr>
<tr>
<td>13</td>
<td>( T_{60,5} \cdot 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 )</td>
<td>Invalid</td>
</tr>
</tbody>
</table>

(Ignition Function)

\[
IGF_j = \frac{TIN \cdot TIW_j}{\alpha + |TIW_j|} \quad (2)
\]

Here \( \alpha \) is kept as a small value close to 0 it’s about 0.0000001. The node which registers the highest ignition function is deemed Wnode i.e.

(Winning Node)

\[ Wnode = \max (IGF_j) \quad (3) \]

In the event of more than one node emerging as the winner owing to the same ignition function value some mechanism such as choosing a node with the smallest index may be devised to break the tie.

The category associated with the winner is the one to which the given input pattern parameters is given by

\begin{align*}
\text{(Weight for Inference)}
WIF_j^{\text{new}} &= \chi(I\Lambda WIF_j^{\text{old}}) + (1-\chi)WIF_j^{\text{old}} \quad (5)
\end{align*}

where \( 0 \leq \chi \leq 1 \)

Once the network has been trained, the inference of patterns, valid or invalid i.e. the categories to which the patterns belong may be easily computed. This is accomplished by passing the input pattern into the preprocessor and then to the input layer.

All the output nodes compute the ignition functions with respect to the input. The winner, node with the highest ignition function, is chosen. The category to which output node belongs is the one to which given input pattern is classified by the network.
4.1.3. DNA Samples

DNA SAMPLE: PERSON 1 [BASE PAIR =32, SEQUENCE =25]
AATGTGTGTTGACCCCTCAAAA
TCTCTCAAATGTGTTTTTACACT
CGTTGGAATGTAATGTTGTTAA
AGTTGCTACCCGGGCTTTTT
AATGTGTCTCT

DNA SAMPLE: PERSON 2 [BASE PAIR =37, SEQUENCE =30]
CAAGTGTGTGGTTACCAAAAATCTC
TCAAATGTGGTGGTTGGGCGTGGT
TAAA TA TGGTAATGTGTTAAAGTGGTG
GTTTGTGGTTAGGGGGGGGCG
GGGGTAATGTGTCTCTGTGGTGGTTAA

(Context Ignition Function)

$$CIF_j = \frac{PPO \cdot WFI_j}{WFI_j}$$

If CIF (1) > CIF (2) the category is valid else if CIF (1) < CIF (2) then category is invalid. For the DNA inputs of fasta format e.g. from person 1 whose category is valid the corresponding seven consecutive components in the DNA sample is chosen as single sequence with base pair thirty two.

Table 3

<table>
<thead>
<tr>
<th>DNA / FASTA Inputs</th>
<th>PREPROCESS SORT ORDER (PPPO)</th>
<th>WEIGHTS FOR INFERENCE</th>
<th>CATEGORY IGNITION</th>
<th>CATEGORY VALID / INVALID</th>
<th>SEPARATOR OUTPUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1, 0.1(AA)</td>
<td>0.1, 0.1, 0.09, 0.09</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9999 0.6428</td>
</tr>
<tr>
<td>0.2, 0.3(TG)</td>
<td>0.2, 0.3, 0.08, 0.07</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9999 0.8571</td>
</tr>
<tr>
<td>0.4, 0.4(CC)</td>
<td>0.4, 0.6, 0.6, 0.04</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9166 0.9285</td>
</tr>
<tr>
<td>0.4, 0.4(CC)</td>
<td>0.4, 0.6, 0.6, 0.04</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9166 0.9285</td>
</tr>
<tr>
<td>0.4, 0.2(CT)</td>
<td>0.4, 0.2, 0.6, 0.08</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9999 0.9166</td>
</tr>
<tr>
<td>0.2, 0.4(TG)</td>
<td>0.2, 0.4, 0.08, 0.06</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9285 0.8571</td>
</tr>
<tr>
<td>0.1, 0.1(AC)</td>
<td>0.1, 0.1, 0.9, 0.09</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9999 0.6428</td>
</tr>
<tr>
<td>0.2, 0.2(CT)</td>
<td>0.2, 0.8, 0.08, 0.02</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.7499 0.6999</td>
</tr>
<tr>
<td>0.2, 0.2(CT)</td>
<td>0.2, 0.8, 0.08, 0.02</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.7499 0.6999</td>
</tr>
<tr>
<td>0.2, 0.1(TA)</td>
<td>0.2, 0.1, 0.08, 0.09</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9999 0.7142</td>
</tr>
<tr>
<td>0.4, 0.3(CG)</td>
<td>0.4, 0.3, 0.06, 0.09</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9999 0.8571</td>
</tr>
<tr>
<td>0.1, 0.1(AC)</td>
<td>0.1, 0.1, 0.9, 0.09</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9999 0.6428</td>
</tr>
<tr>
<td>0.2, 0.1(TA)</td>
<td>0.2, 0.1, 0.08, 0.09</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9999 0.7142</td>
</tr>
<tr>
<td>0.3, 0.4(GG)</td>
<td>0.3, 0.4, 0.07, 0.06</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9999 0.9285</td>
</tr>
<tr>
<td>0.3, 0.3(GG)</td>
<td>0.3, 0.7, 0.7, 0.03</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.8333 0.9999</td>
</tr>
<tr>
<td>0.3, 0.3(GG)</td>
<td>0.3, 0.7, 0.7, 0.03</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.8333 0.9999</td>
</tr>
<tr>
<td>0.3, 0.2(CT)</td>
<td>0.3, 0.2, 0.7, 0.08</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9999 0.7857</td>
</tr>
<tr>
<td>0.1, 0.1(AC)</td>
<td>0.1, 0.1, 0.09, 0.09</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9999 0.6428</td>
</tr>
<tr>
<td>0.4, 0.2(CT)</td>
<td>0.4, 0.2, 0.06, 0.08</td>
<td>0.1, 0.1, 0.5, 0.05</td>
<td>0.2, 0.5, 0.5, 0.2</td>
<td>~</td>
<td>0.9999 0.9166</td>
</tr>
</tbody>
</table>

Valid Sequences (V)

$$V_{p, s, k} = V_{seq} \quad p, s, 1, V_{seq} \quad p, s, 2, \ldots, \quad V_{seq} \quad p, s, k$$

The DNA inputs of fasta format whose category is invalid, the input is considered as a sequence with two consecutive components. To generate the unique identification code for each individual the set of valid sequences is used.
Table 4

<table>
<thead>
<tr>
<th>S.No</th>
<th>p(Person)</th>
<th>s(Sequence)</th>
<th>Discriminator Inputs (V_{p.s,k})</th>
<th>Vseq_{p.s,k} (k=1)</th>
<th>Vseq_{p.s,k} (k=2)</th>
<th>Vseq_{p.s,k} (k=3)</th>
<th>Vseq_{p.s,k} (k=4)</th>
<th>Vseq_{p.s,k} (k=5)</th>
<th>Vseq_{p.s,k} (k=6)</th>
<th>Vseq_{p.s,k} (k=7)</th>
<th>Unique Identification number (UN_{s})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>V_{1,1,k}</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.182464</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>V_{1,2,k}</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>0.442375</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>V_{1,3,k}</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.672457</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>V_{1,4,k}</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>0.1</td>
<td>0.672577</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>5</td>
<td>V_{1,5,k}</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.182464</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>6</td>
<td>V_{1,6,k}</td>
<td>0.2</td>
<td>0.1</td>
<td>0.4</td>
<td>0.1</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>0.675453</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>7</td>
<td>V_{1,7,k}</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.670144</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>8</td>
<td>V_{1,8,k}</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td>0.151905</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>9</td>
<td>V_{1,9,k}</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.182464</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>10</td>
<td>V_{1,10,k}</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.236024</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>11</td>
<td>V_{1,11,k}</td>
<td>0.3</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.731645</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>12</td>
<td>V_{1,12,k}</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.710014</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>13</td>
<td>V_{1,13,k}</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.182464</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>14</td>
<td>V_{1,14,k}</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.678400</td>
</tr>
</tbody>
</table>

Figure 5: The Chart Showing the Unique Identification Number for the Sample of PERSON 1 i.e. 0.182464

Where \( p, s = 1 \) to \( a \) and \( k = 1 \) to \( 7 \)

The valid separator output are fed to the discriminator where the discriminator output defined by

\[
D_{p,s} = \sum_{k=1}^{7} k(V_{seq_{p,s,k}})^k
\]

generates the unique identification number.
5. CONCLUSION

In this paper the Bio-Neural Pattern Recognition Tool is used to classify the sequences that are used to identify a unique number from the given human DNA sample which actually includes almost five nucleotide basis of adenine, guanine, cytosine, thymine and uracil, which are represented by five fuzzy values respectively. In the field of Bio-Informatics it would be inevitable in future to deal with numerous nucleotide compositions, their classifications in case of protein sequencing, mutation and so on, for which the above tool can be used extensively since it uses neural and fuzzy approach with a long range of fuzzy values between 0 to 1 instead of a minimal range of real values, which overcomes complications in the generation of algorithms using conventional mathematical techniques that involves more iterations in it.

References


