DNA Pattern Analysis

using Finite Automata (FA), Mealy and Moore Machines

- **JOURNALIST(S):** **Murtaza Mehdi**

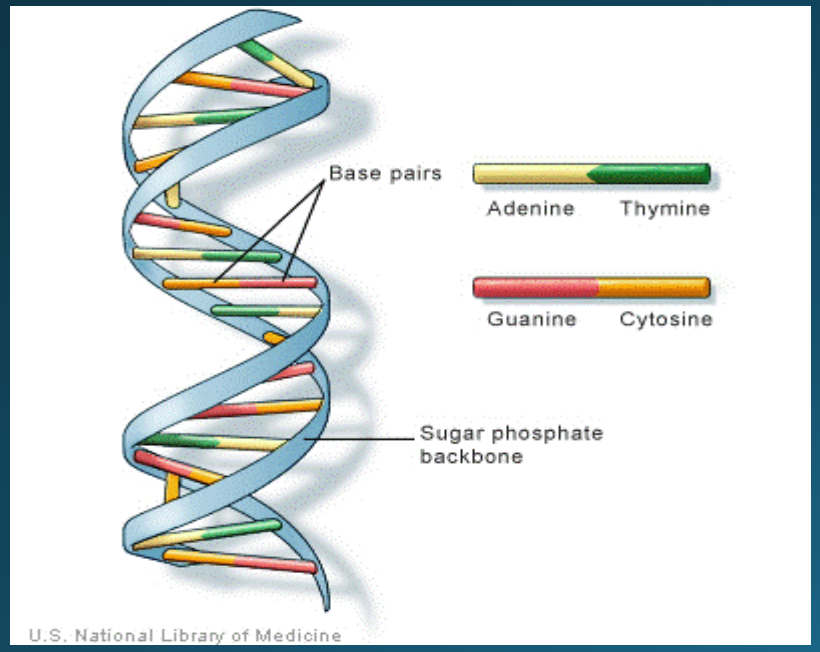
  (Department of Software Engineering, Foundation University Rawalpindi Campus, Pakistan)

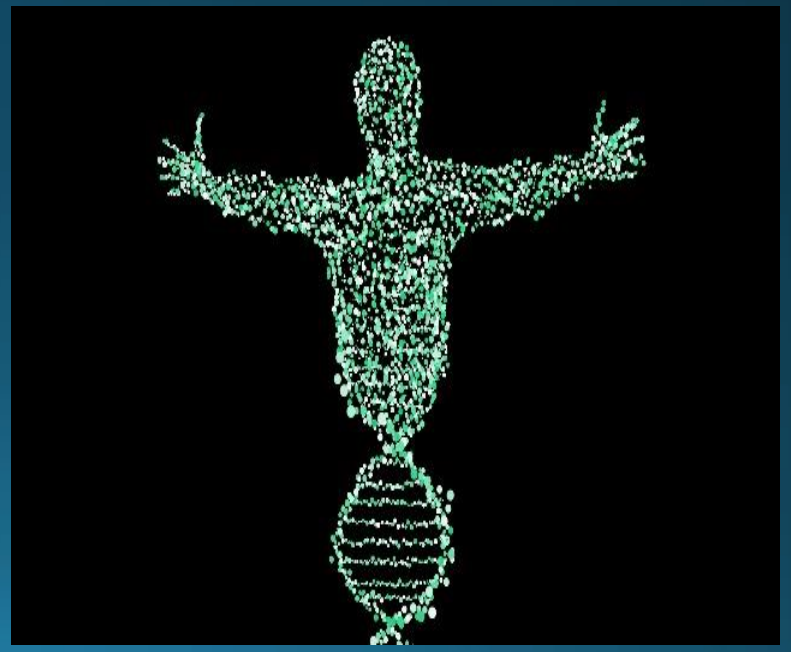
  Aihab Khan

  (Department of Computer Science, Iqra University, Islamabad Campus, Pakistan)

- **YEAR PUBLISHED:** 8, August 2016
- **JOURANL NAME:** International Journal of Computer Science and Information Security (IJCSIS), Vol. 14

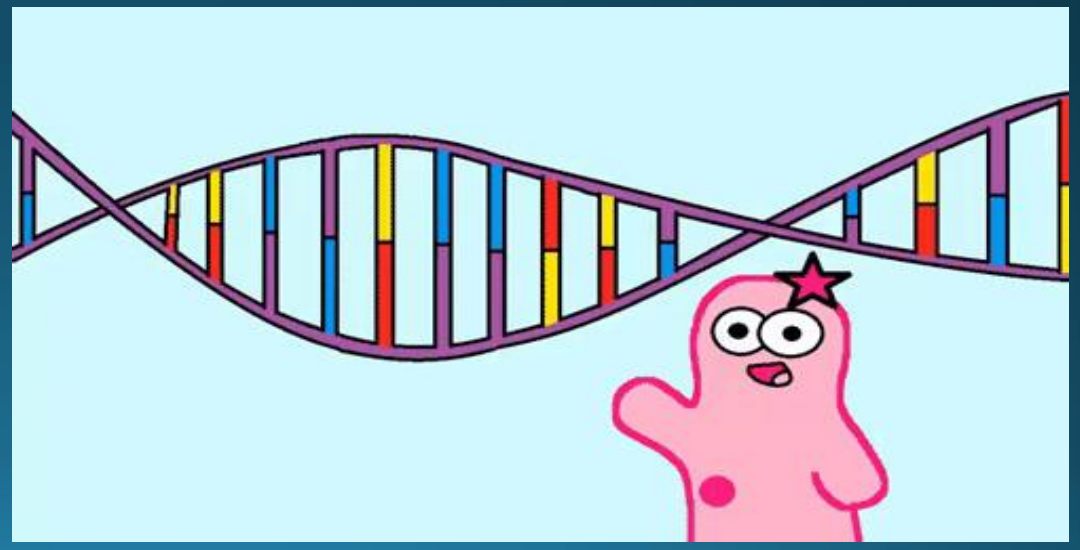
- **PURPOSE:** The **purpose** of **using FA** for **DNA pattern analysis** is that any change, alteration or duplication **in** gene or **in** genetic information can be detected.
- **NEED:** **Your DNA** and genes determine certain traits such as height and eye colour, as well as how **our** body functions. For example, **your** genes can determine whether you're likely to **have** food intolerances, vitamin deficiencies, how you metabolise different parts of **your** diet, and aspects of **your** skin's reaction to sunlight.



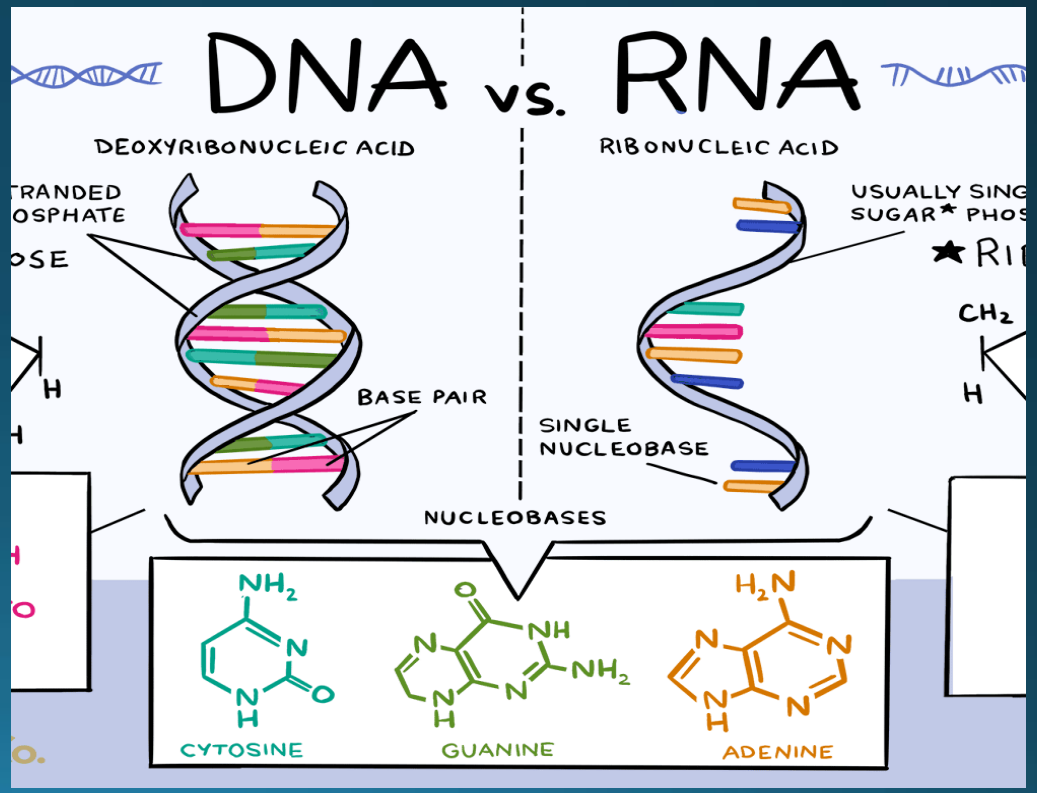


What is DNA?

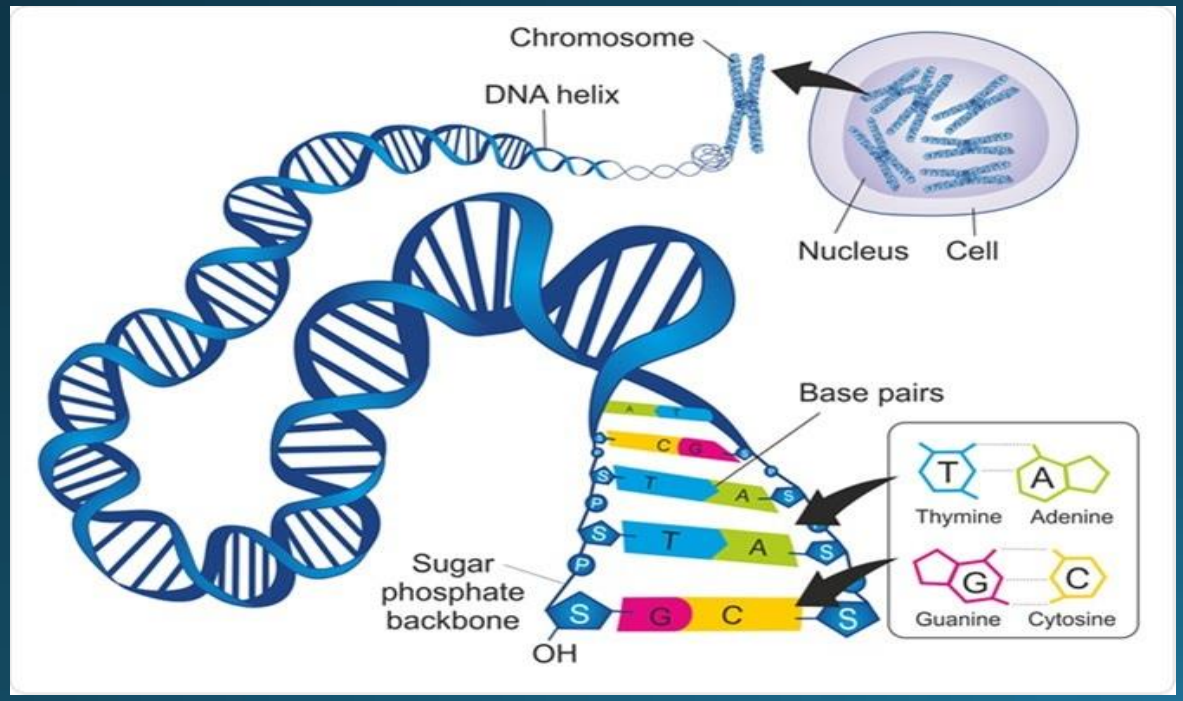
- Deoxyribonucleic Acid(DNA) is a series of genes made by the mixture of nucleotides.
- The adjustment in the arrangement of these nucleotides can change the hereditary data that cause numerous issues in living creatures.

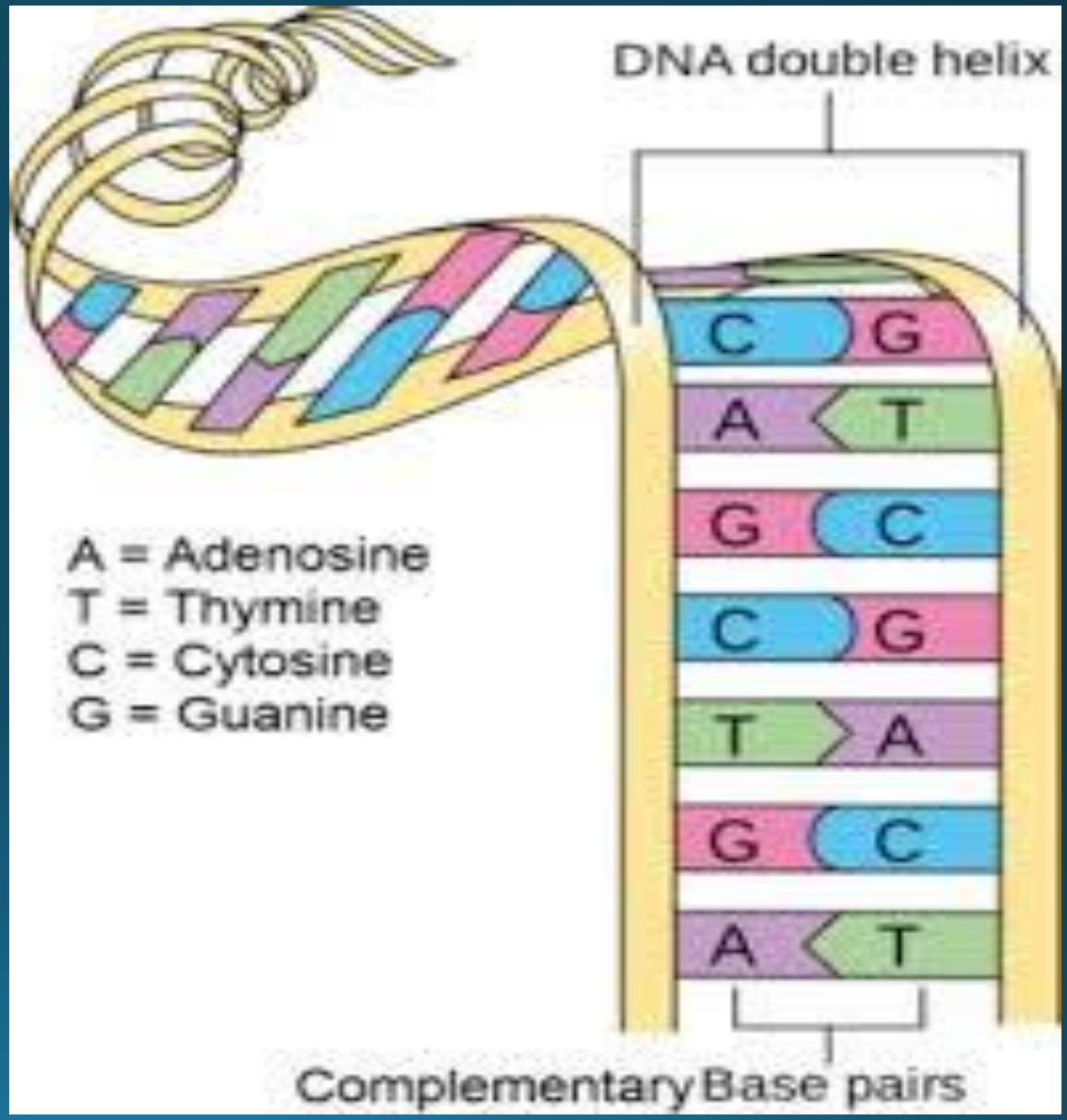


- It is useful to find out such change in nucleotides to prevent such type of disorders and abnormalities.
- In genetic code of DNA, nucleotides in different sequences are used to form genetic information.
- These nucleotides are also called codons and combination of codons which helps in protein.
- For DNA pattern conversion there are **four nucleotides** that helps to understand pattern and transcribe into RNA pattern and then this RNA pattern translates into protein for finding any problem in pattern.



- The four important nucleotides are:
  - ➢ Cytosine (C),
  - ➢ Adenine (A),
  - ➢ Thymine (T) and
  - ➢ Guanine (G).



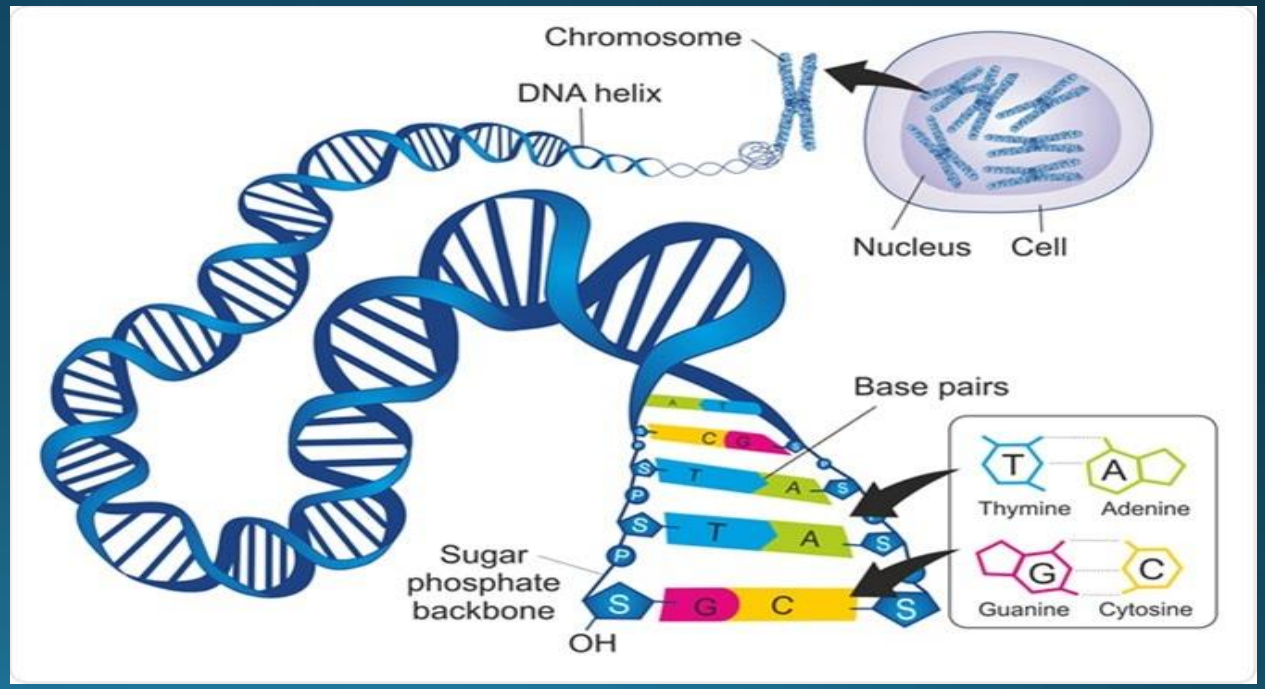


What is the role of Computer Science to detect Pattern of DNA?

- There are methods in theoretical computer science such as **Finite Automata, Mealy machine and Moore machine** to examine patterns of DNA and analyze any change in nucleotides to prevent various genetic disorders. In the present study, **these three methods were combined to evaluate their performance**.
- It was found that the Mealy machine is much better as compared to others because it has less states, and react faster to inputs.
- We use NFA and DFA to analyze DNA pattern.
- The DNA is converted into a pre-mRNA then pre-mRNA is converted into mRNA and then mRNA converted into RNA which finally transformed into protein in final state. (by using enzymes)

Cepheid
GeneXpert Omni

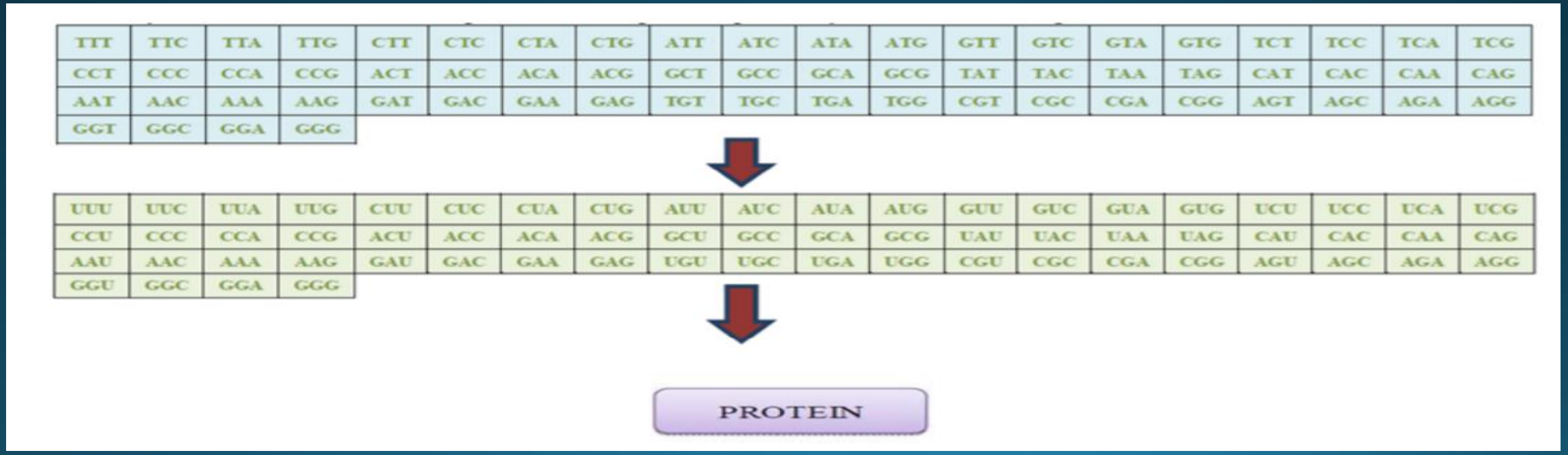
DNAscan


Chromosome
DNA helix
Nucleus
Cell
Base pairs
Sugar phosphate backbone
OH
Thymine
Adenine
Guanine
Cytosine

Methodology:

DNA pattern:

- In DNA consists of combinations of four Codons which are known as Cytosine (C), Adenine (A), Thymine (T) and Guanine (G).
- Transition of these codons into protein is feature in below figure:



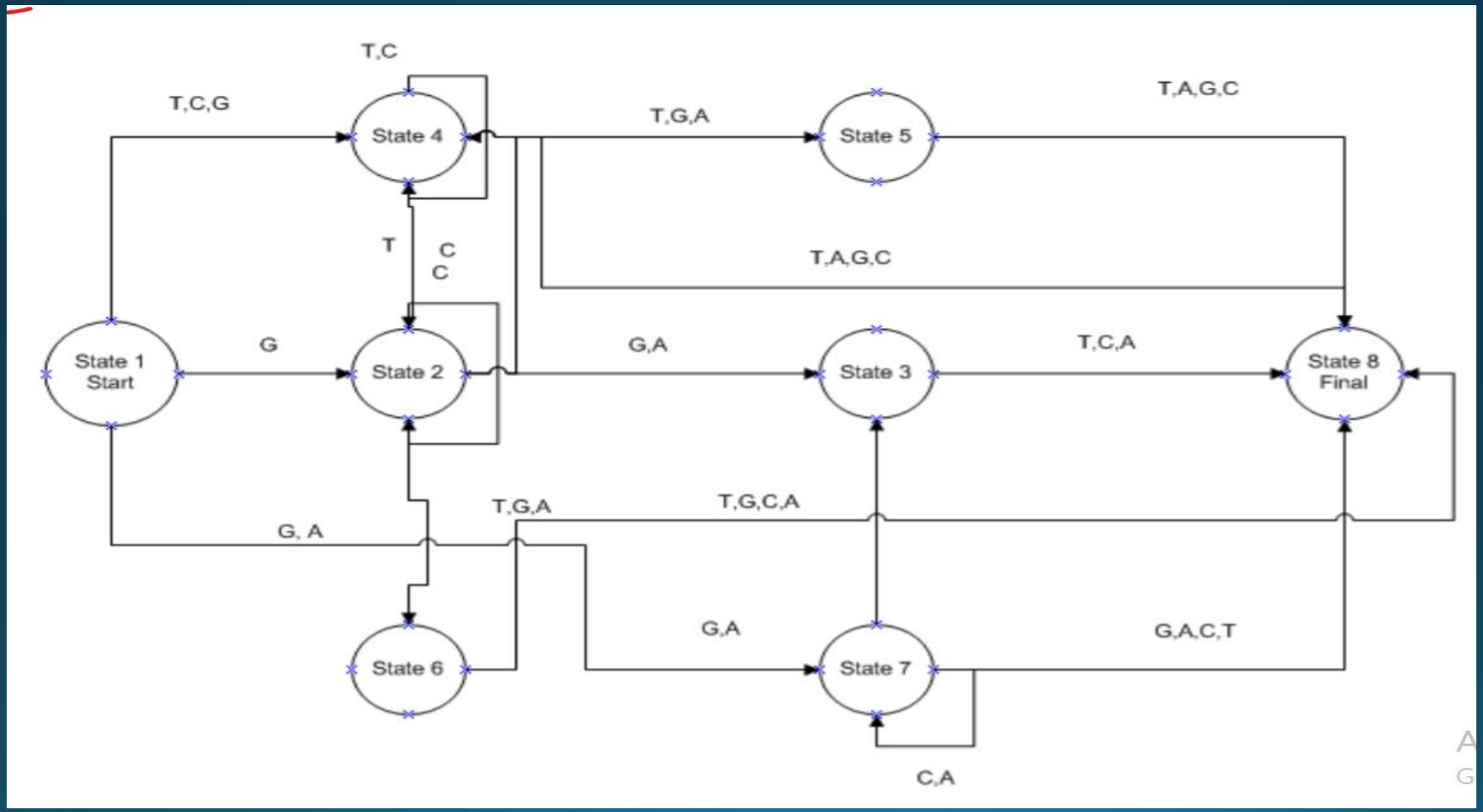
1) DNA Pattern Analysis with Finite Automata:

- In the proposed model, **eight states** were used to analyze DNA normal and abnormal patterns using JFLAP.

$Q = \{ 1,2,3,4,5,6,7,8 \}$

$\Sigma = \{ A,T,C,G \}$

- The 8 states of DNA pattern, in which on every state were checked for the inputs of DNA general pattern, if an input was accepted by given machine it was to be regarded as a normal pattern otherwise some abnormality in specific pattern was be assumed.


T,C
T,C,G
State 4
T,G,A
State 5
T,A,G,C
T
C
C
T,A,G,C
State 1
Start
G
State 2
G,A
State 3
T,C,A
State 8
Final
T,G,A
T,G,C,A
G, A
G,A
G,A,C,T
State 6
State 7
C,A

- Finite Automata Experiments and Results:

DFA

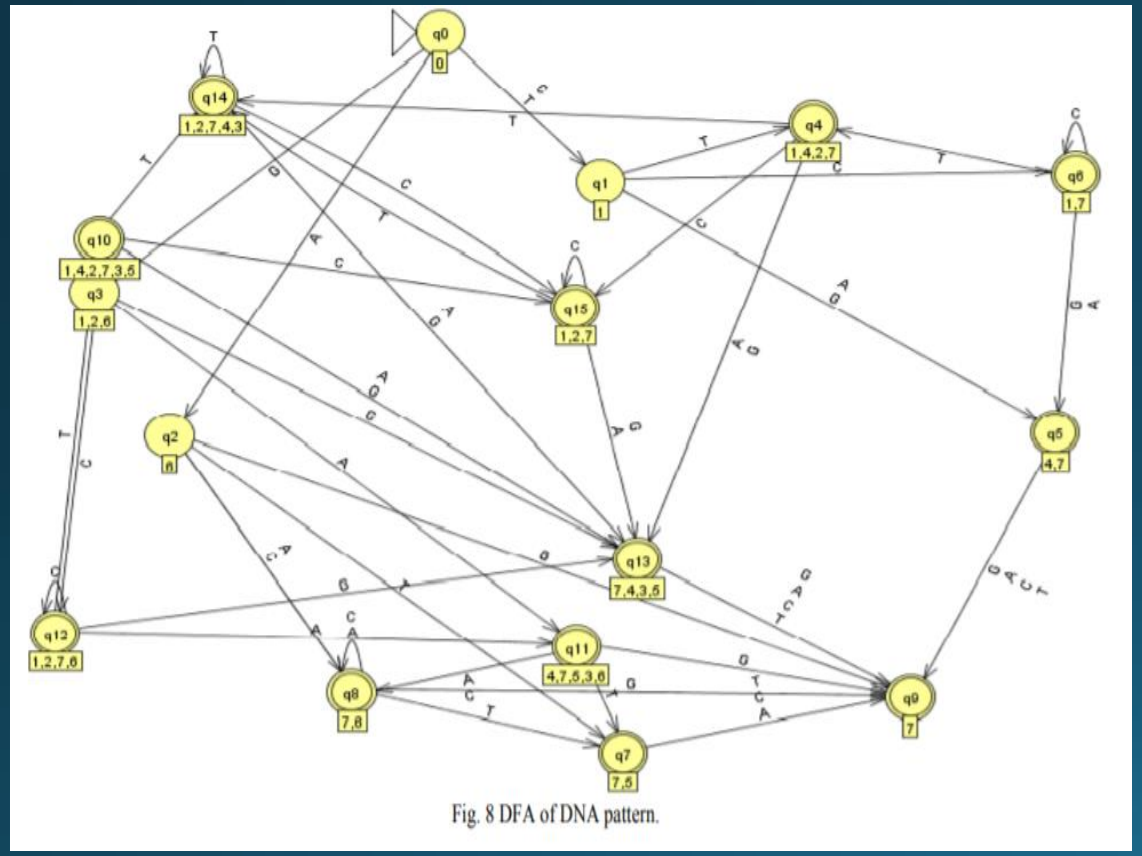


Fig. 8 DFA of DNA pattern.

NFA

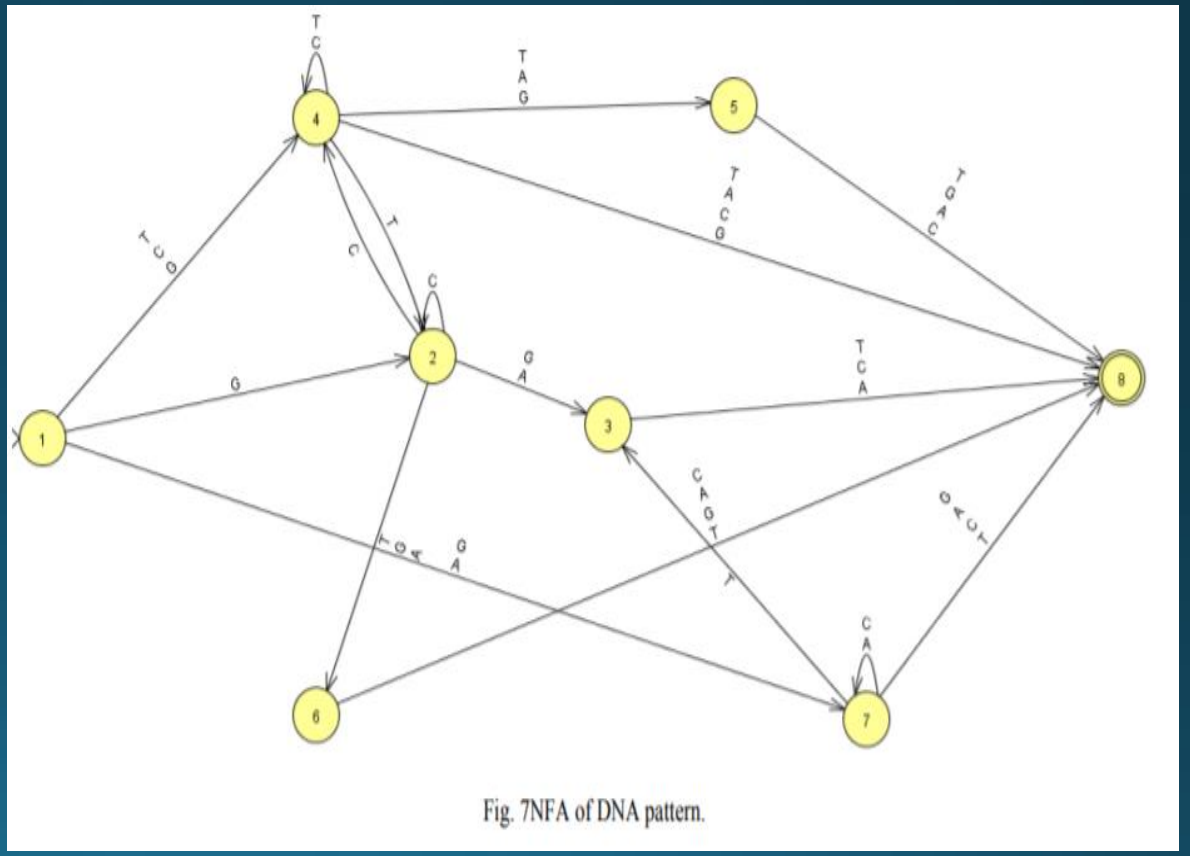


Fig. 7NFA of DNA pattern.

Results for DFA & NFA

Input	Result
TTT	Accept
TCG	Accept
TAG	Accept
GTA	Accept
TCG	Accept
AAA	Accept
GGT	Accept
GGA	Accept
TCT	Accept
ACA	Accept
GGT	Accept
GGC	Accept
GGA	Accept
GGG	Accept
CGC	Accept
ATC	Accept
ATCG	Reject
GGTA	Reject

2) DNA Pattern Analysis with Mealy Machines:

- Random inputs were selected and checked for Mealy machines acceptance. In case the machine accepted the input on every stage, the DNA was normal otherwise it was abnormal.
- Input (A, G, C, T)
- There are five States.

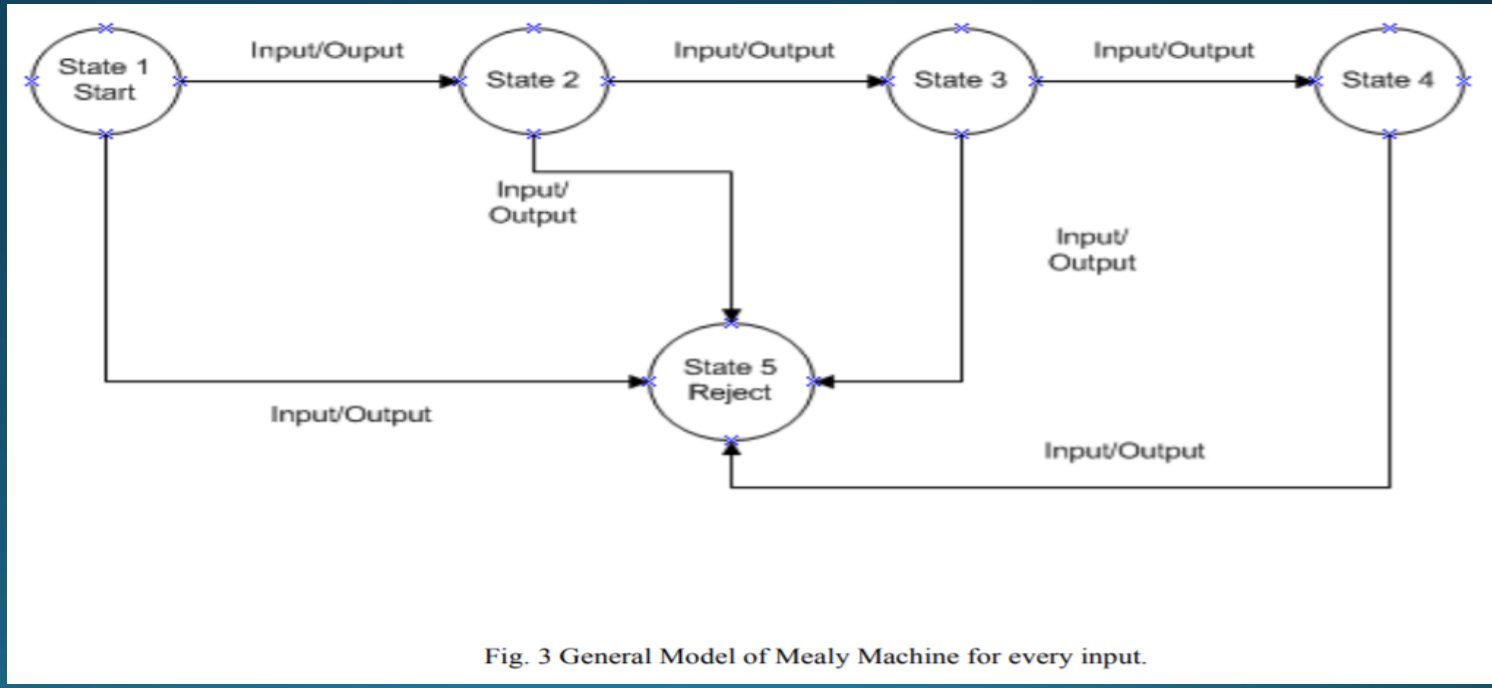


Fig. 3 General Model of Mealy Machine for every input.

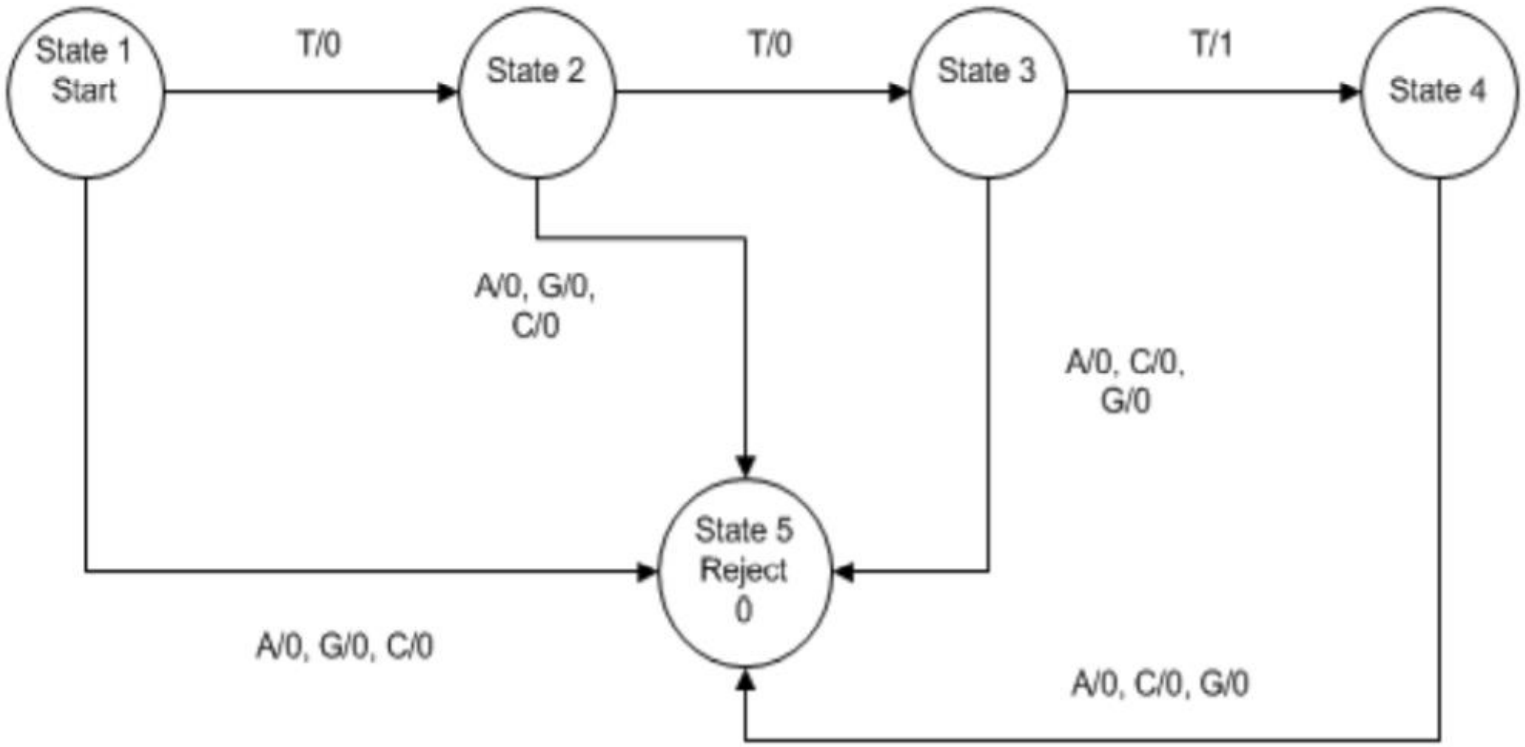


Fig. 4 Experiment Model of Mealy Machine for specific input.

- DNA pattern of Mealy Machine on some inputs (TTT, ATG, TAG, and CGA) are given as under: TTT The following figure accepts DNA pattern for 'TTT'.

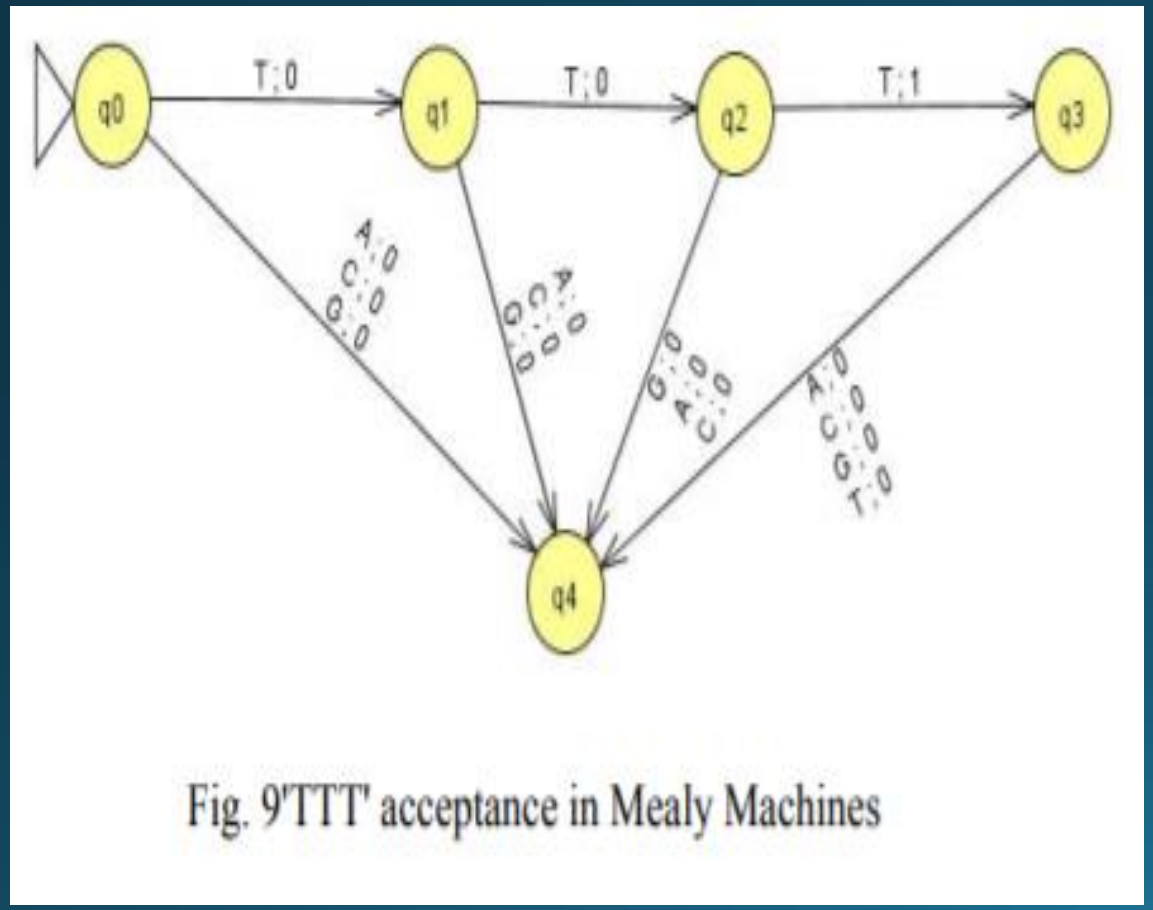


Fig. 9'TTT' acceptance in Mealy Machines

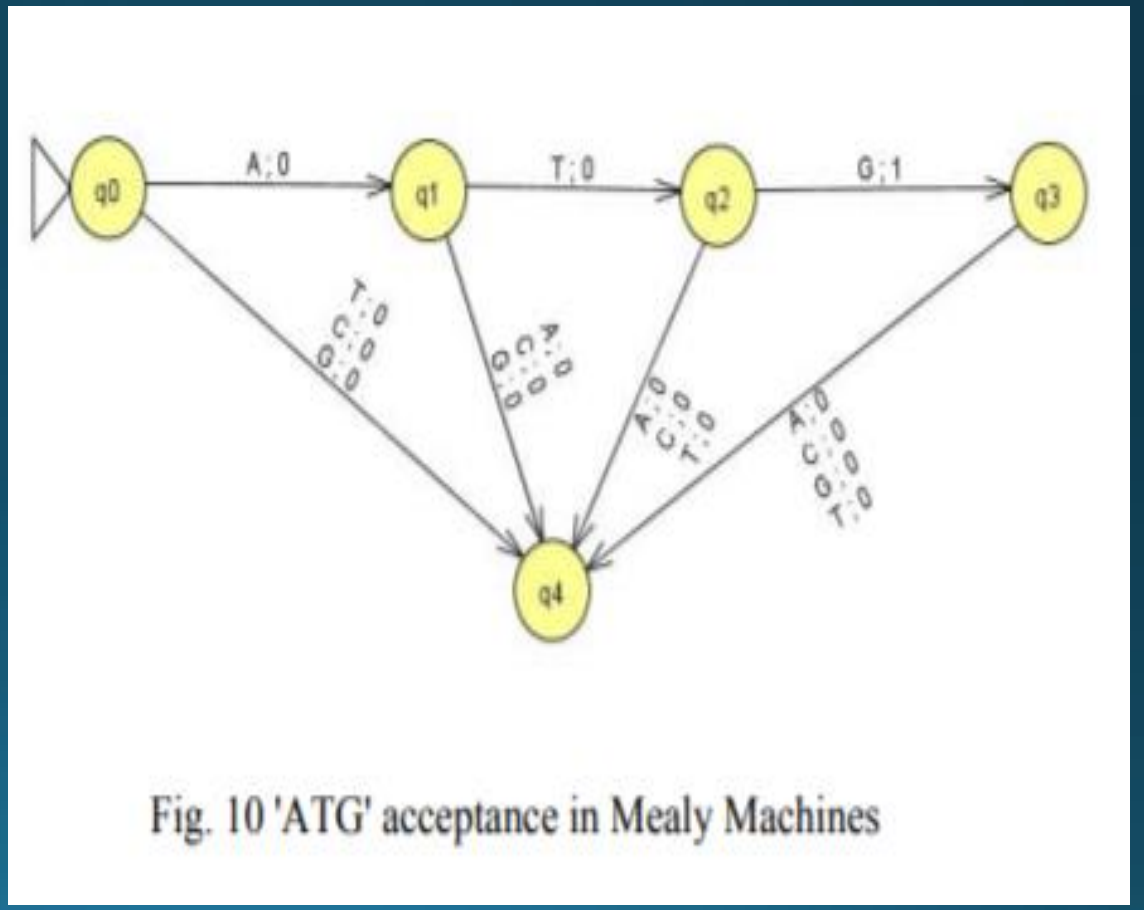


Fig. 10 'ATG' acceptance in Mealy Machines

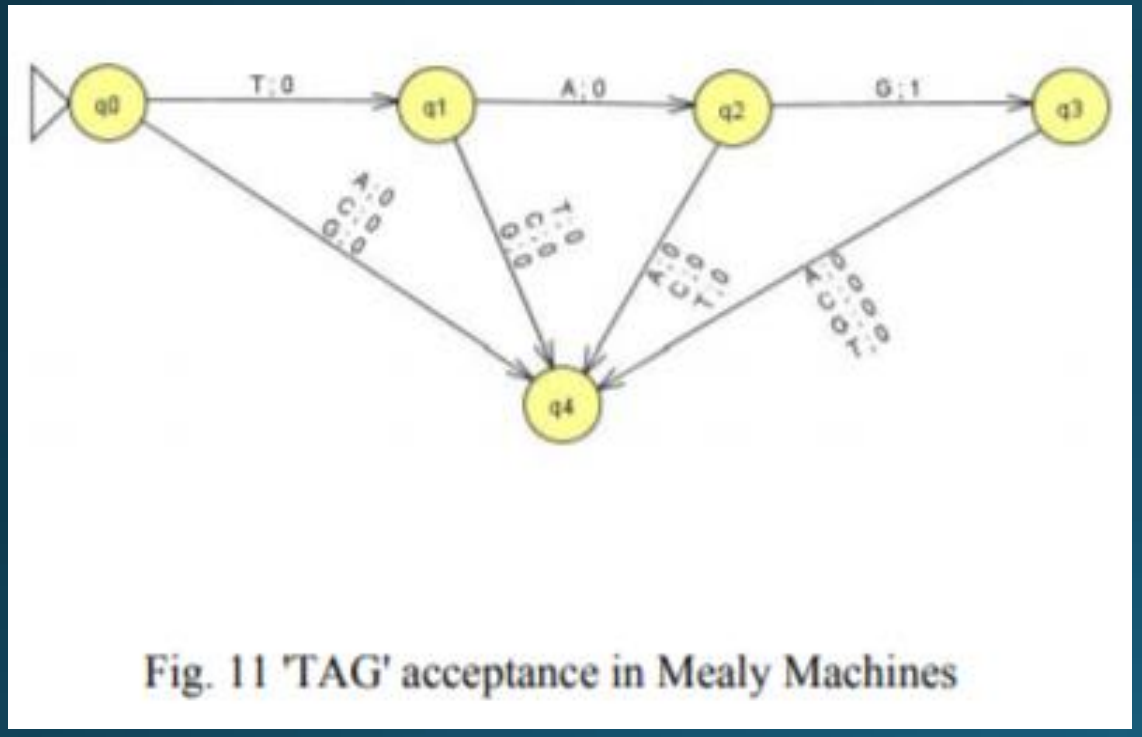


Fig. 11 'TAG' acceptance in Mealy Machines

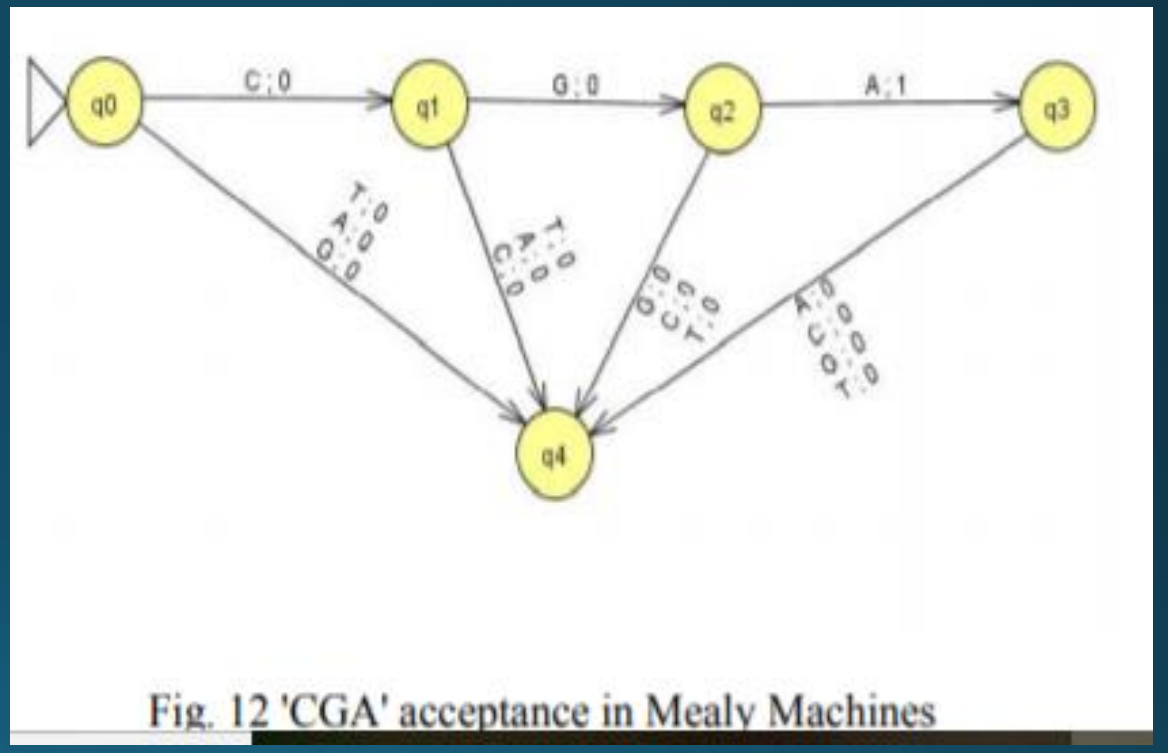


Fig. 12 'CGA' acceptance in Mealy Machines

TTT result:

The following table shows result for acceptance of DNA pattern for ‘TTT’.

Input	Result
TTT	001
TGA	00
AGT	0

3) DNA Pattern Analysis with Moore Machines:

- Random inputs were selected and checked for every input that Moore machines acceptance . It was inferred that in case the machine accepted the input on every stage, the DNA will be normal otherwise it was proven to be abnormal. There is a general model through which various inputs can be checked for the acceptance and rejection by the machine.
- Input (A, G, C, T)
- There are five States.

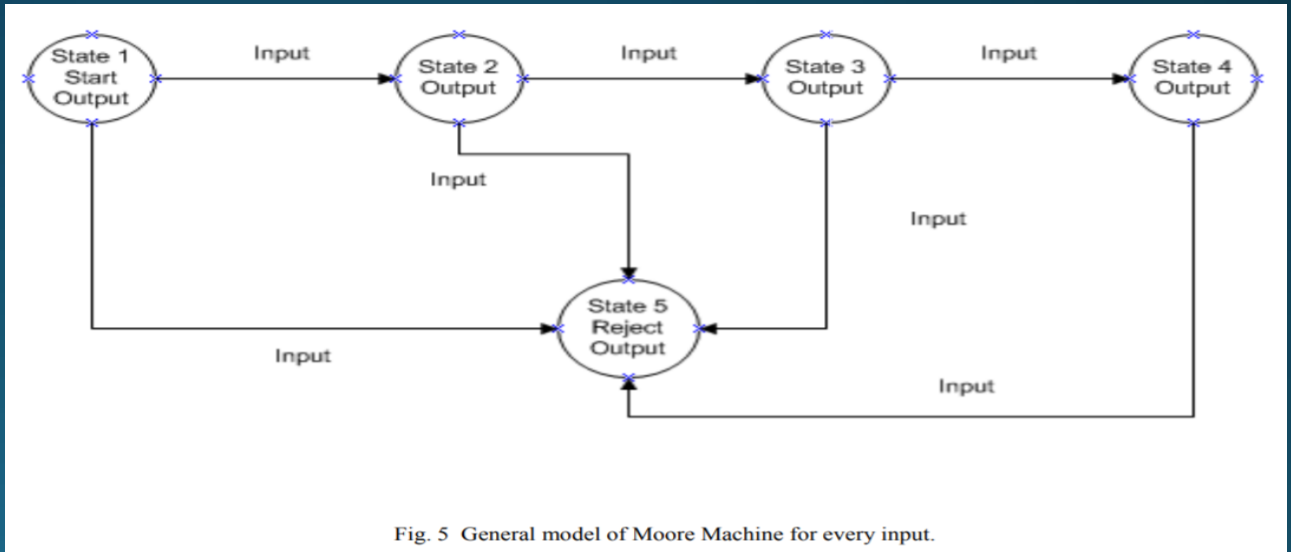


Fig. 5 General model of Moore Machine for every input.

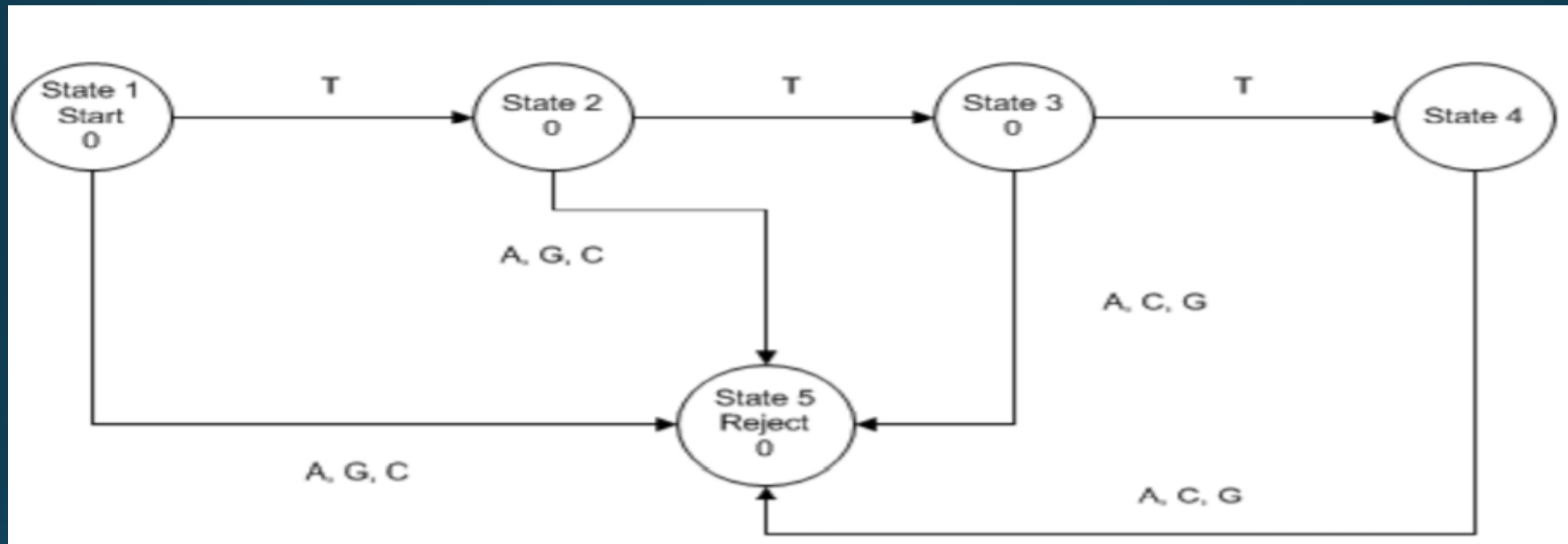


Fig. 6 Experiment model of Moore Machine for specific input.

- DNA pattern of Mealy Machine on some inputs (TTT, ATG, TAG, and CGA)

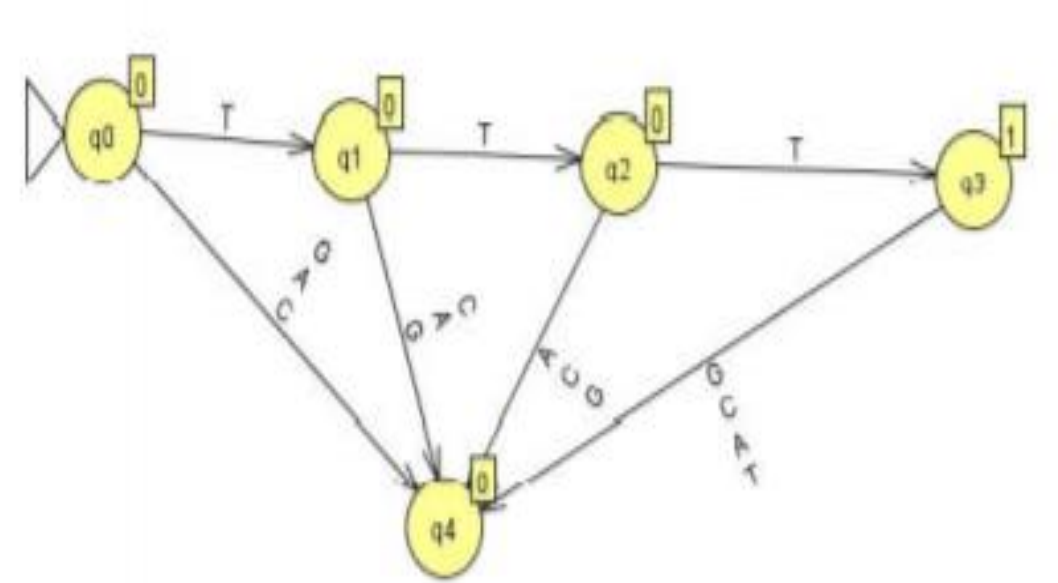


Fig. 13 'TTT' acceptance in Moore Machines

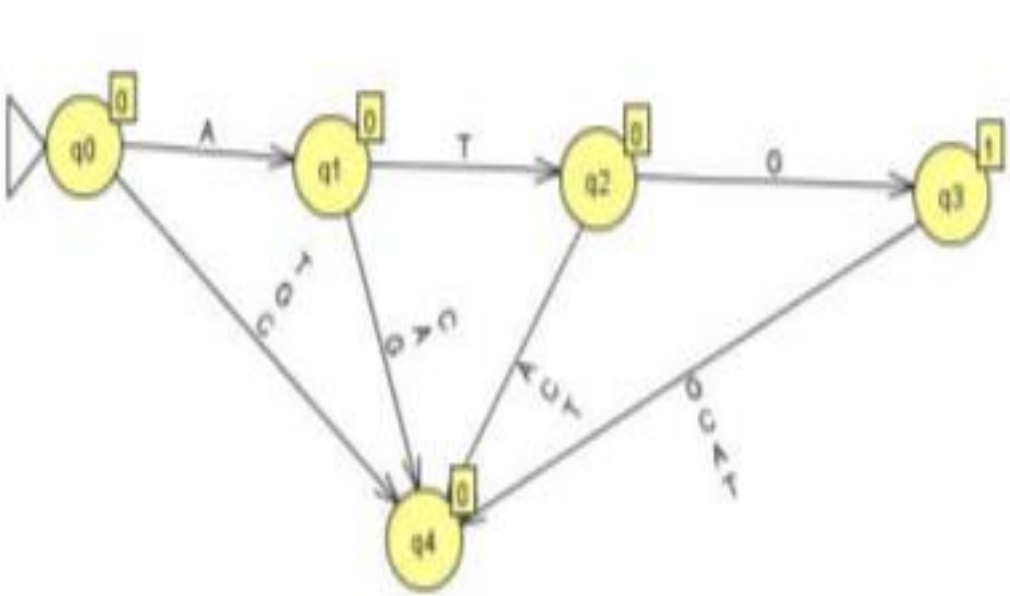


Fig. 14 'ATG' acceptance in Moore Machines

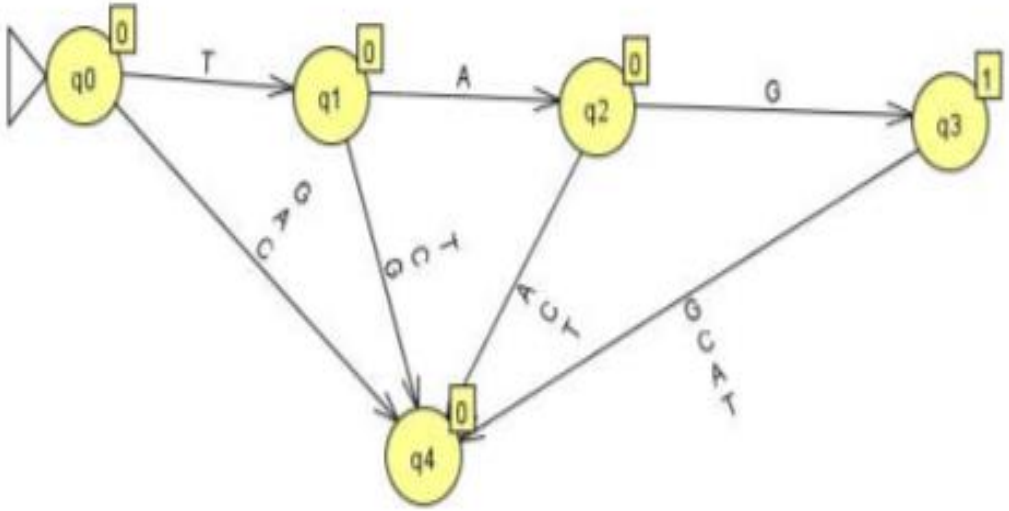


Fig. 15 'TAG' acceptance in Moore Machines

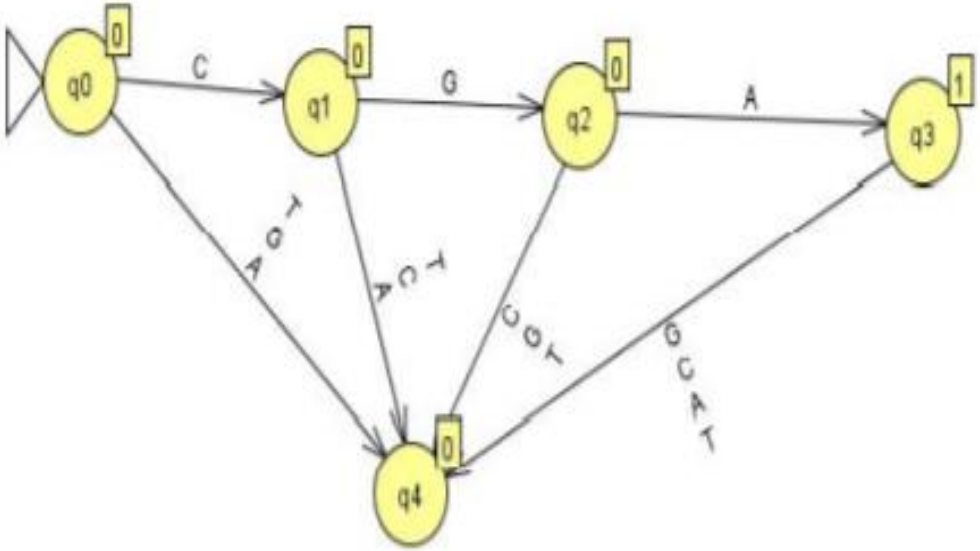


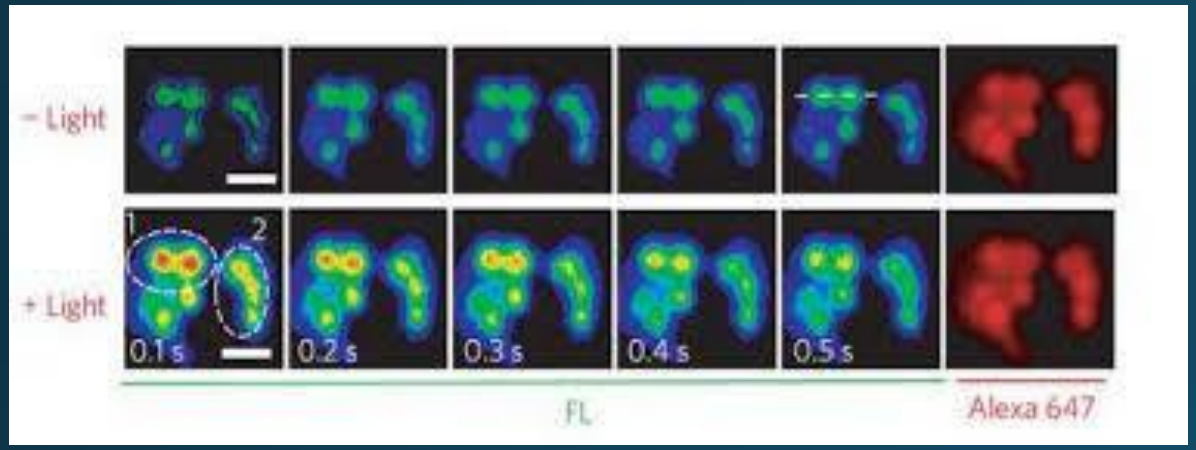
Fig 16 'CGA' acceptance in Moore Machines

CGA result:

The following table shows result for acceptance of DNA pattern for ‘CGA’.

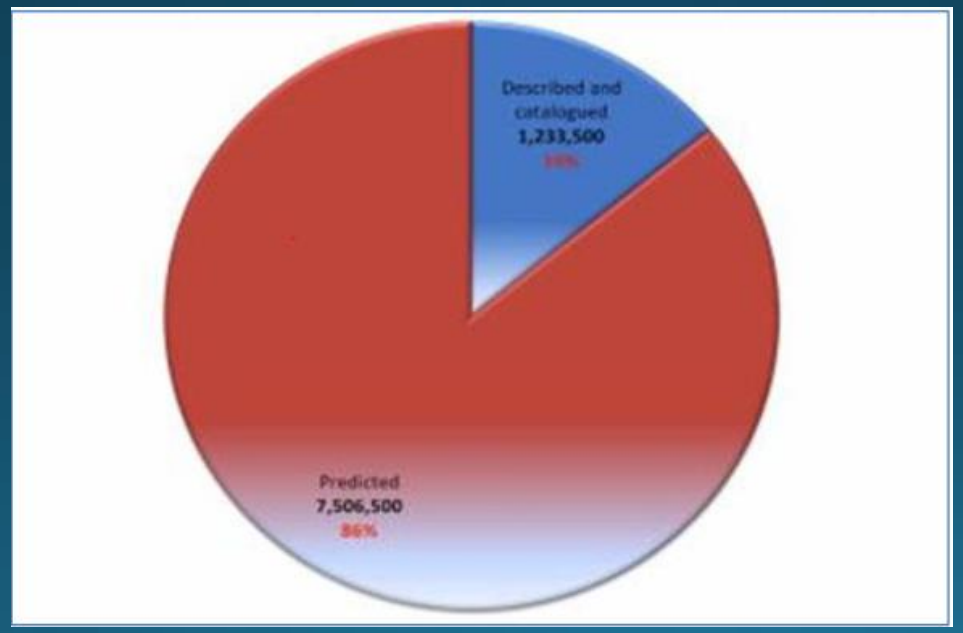
Input	Result
TTT	00
TGA	00
AGT	00
CGA	0001

Technology have vital and significance role in science or you can say vice versa.

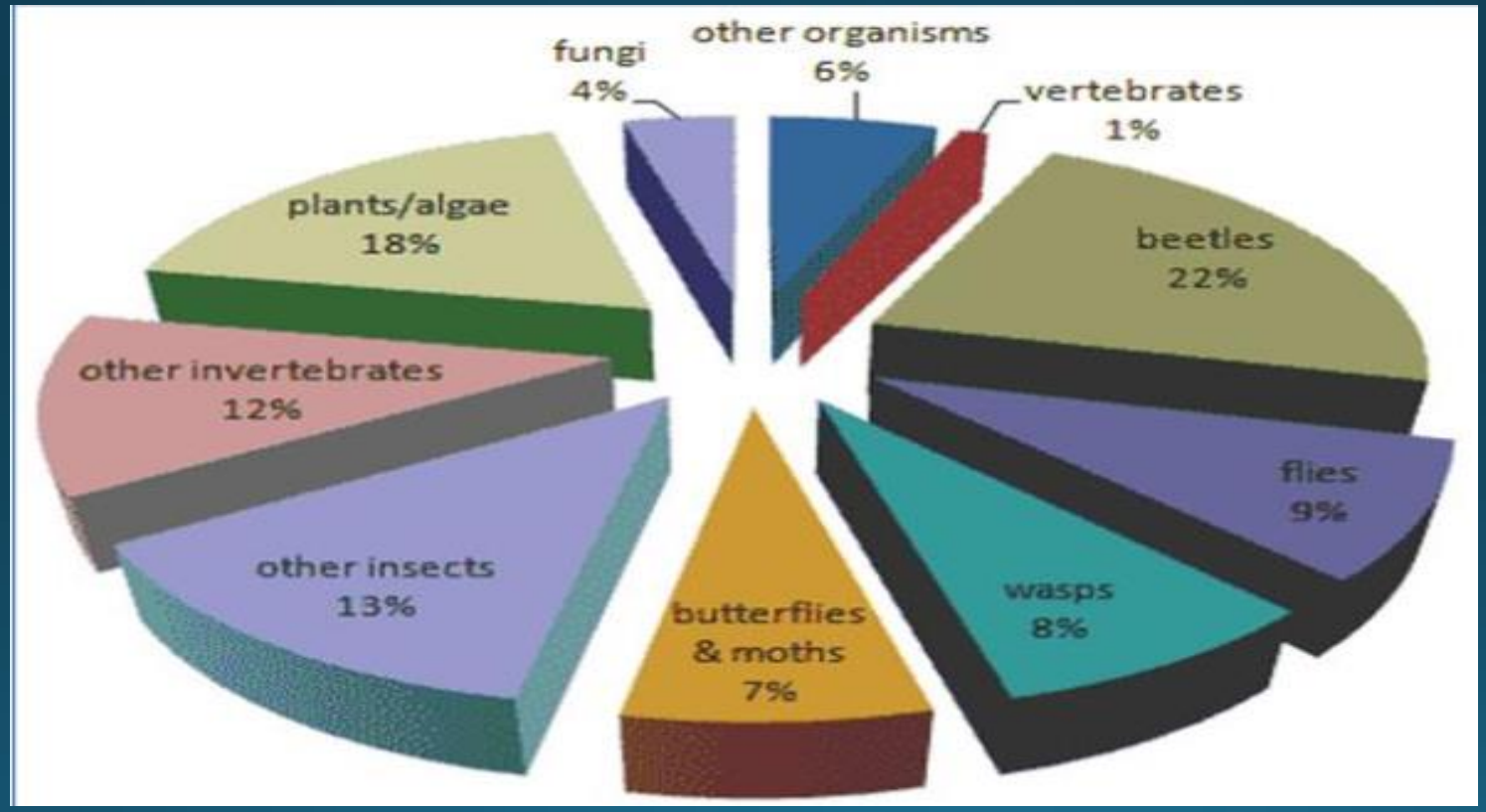


DNA samples of living organisms

- There are approximately 8.7 million species of species on our planet out of which 6.5 million are from land and the remaining from the seas.



- As shown in the previous figure, only 1.8million species have been categorized and known to mankind. This clearly states that around 75-90% of them are yet to be discovered.

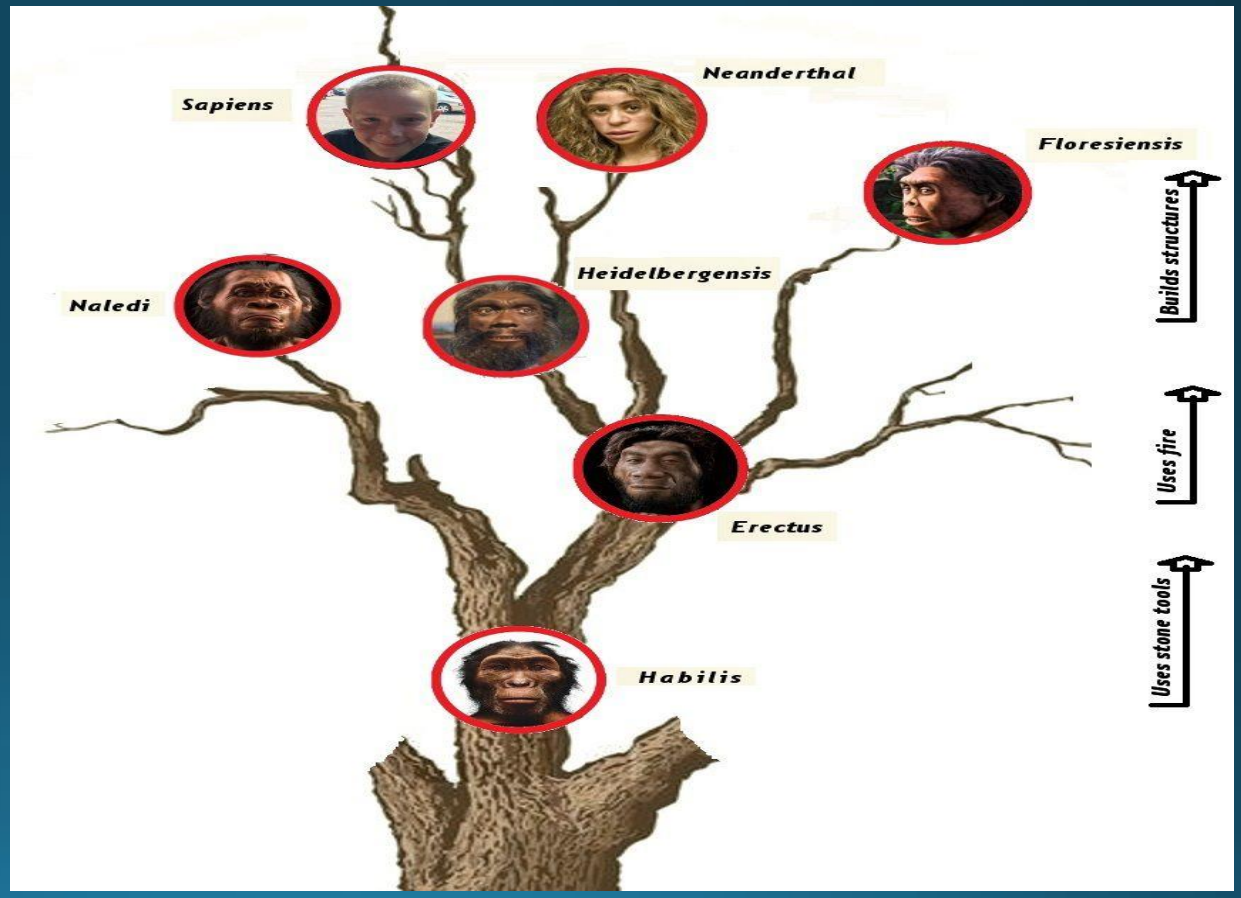


How a DNA revolution has decoded the origins of our humanity

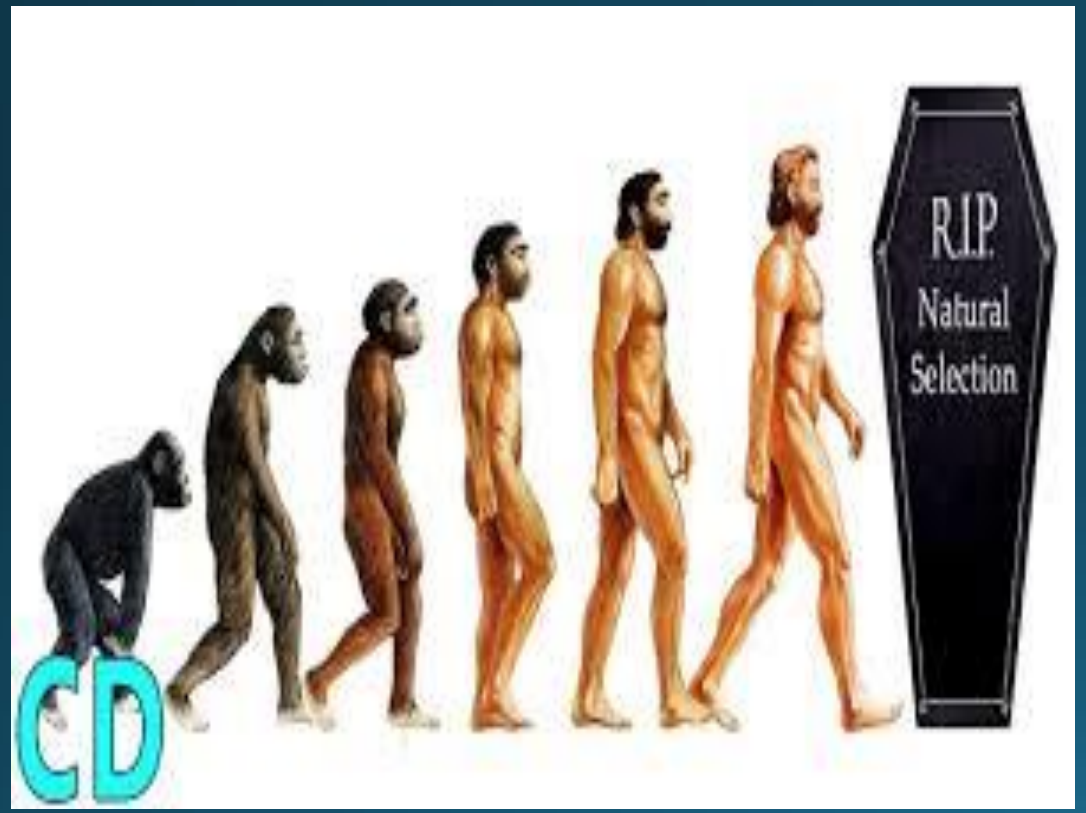
When errors do occur during copying, mutations arise. Some mutations are beneficial, and some are not. If the mutations occur in sex cells, they can also be passed from parents to offspring. The existence of random mutations is essential for evolution theory. Populations will naturally vary; some individuals may have certain mutations while others do not. Those with beneficial mutations may be more likely to survive and produce offspring, passing their mutation to some of their offspring.

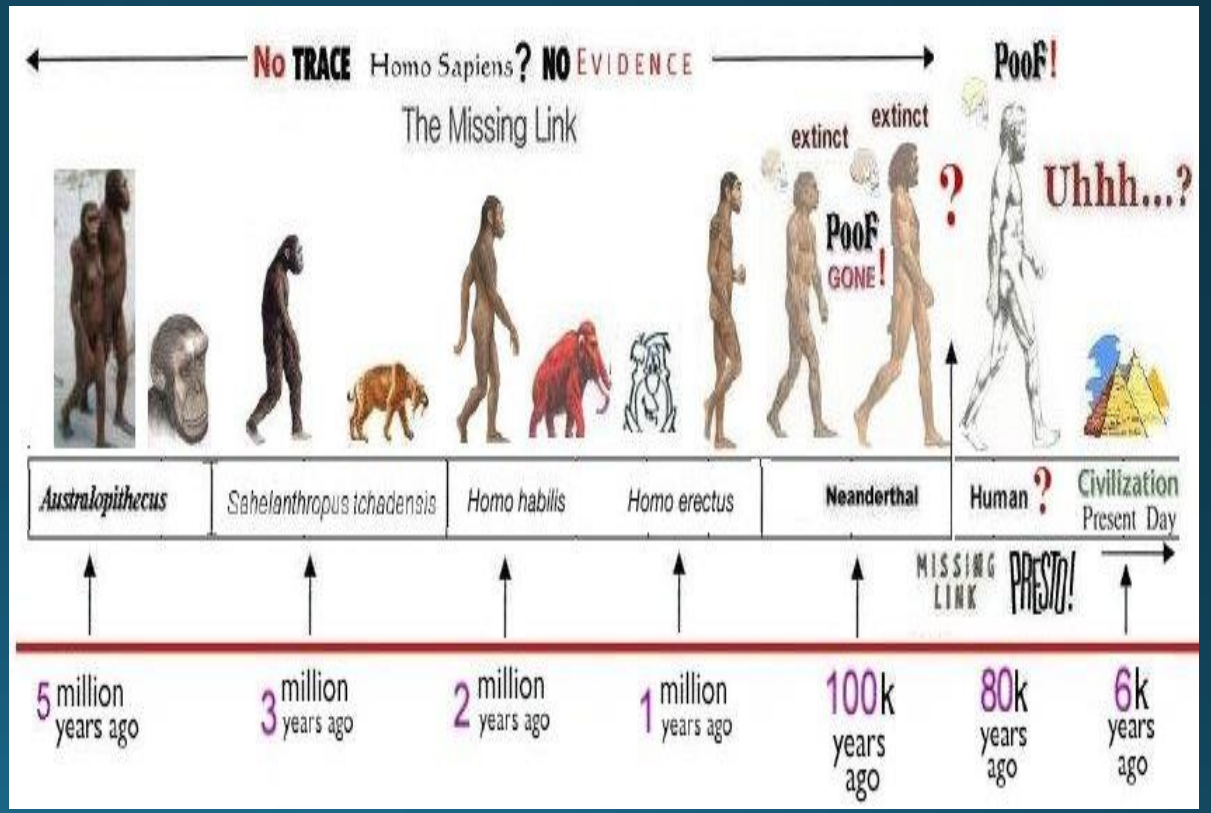
- Extract from the book “Sapiens-a brief History of Humankind”(recommended book):

  “Earlier traditions usually formulated their theories in terms of stories. Modern Science uses mathematics.”



- Some illustrations of human evolution are under:





- Genetics play much vital role in order to track the human history.
- Crack the history of human migration.

Ethnicity Regions

Native American—North, Central, South	62%
Central & Southern Mexico	
Western Jalisco	
Spain	21%
France	6%
Italy	2%
Cameroon, Congo, & Southern Bantu Peoples	2%
Ivory Coast/Ghana	2%
Mali	2%
Basque	1%
Philippines	1%
Senegal	1%

CONCLUSION

- Mealy machine is much better than Moore machine.
- The results show the performance of Mealy machines is better for analyzing the DNA pattern because the Output depends both upon present state and present input.
- Mealy machines react faster to inputs. In future we work to analyze pattern of DNA through other computational models compare with these machines for performance.
- For mathematical results, theory of automata proves to be vital importance.

FUTURE WORK

- Although 1.8 million species are discovered today, all their DNA nucleotides are not easily accessible to study the differences and the similarities between these organisms.
- Also, DNA can be represented in 3D structures depending on the behavioral patterns of proteins in the amino acids. This can be achieved in future research.

"DNA Pattern Analysis using FA, Mealy and Moore Machines:"

Case study Research

- Course No: CSSE-413
- Presented by: Wajiha Hanif Arain (B18158064)
  Yumna Noor (B18158066)
  Muhammad Zaeem khan (B18158042)
- Faculty: BSSE- 3rd semester
- Presented to: Ma'am Farheen Faisal