

Tracing Missing Children by DNA Matching Using Bayes Classification Algorithm

Mr.L.Jaga.Jeevan.Rao¹, Pushpa Kanda², Yanaparathi Poornima³, Relangi Ramya⁴, Reyvi Likhita⁵

Assistant Professor, Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women Visakhapatnam, India¹

Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women Visakhapatnam, India²⁻⁵

Abstract- Talking about the present scenario, we all are aware how unsafe our society is for children and women. There has been an alarming increase in the child missing reports registered in different parts of the world. According to a report published by The Indian Express, around 22,567 children went missing out of which 12,999 were traced back in 2019 and according to Delhi Police PCR around 17-12(2019-2020) kids go missing everyday which is very traumatic for the loved ones and the relatives of the children. It shows how many loopholes are there in our law enforcement agencies. There can be varied reasons for missing children like abduction or kidnapping by someone, human trafficking, child willingly leaving his/her home, etc. To find a match between specific parent and child DNA we use Bayes Classification Repeats.

I. INTRODUCTION

Today there are a lot of missing cases seen around the world. These missing children who are traced are placed in child line centers. But the major challenge arises in identifying the identity of these children. In order to provide a solution to this problem, DNA matching is used.

DNA matching is the process of sequencing DNA and comparing it to the DNA of other people i.e., children in a database when a parent with significant amount of DNA that is identical to child is found, that can indicate that both of them are related.

The algorithm that we use to match DNA is Bayes classification algorithm and short Tandem repeats. It is used to find the repeated sequence of DNA in both parent and child. Later Bayes Classification finds the percentage of DNA that is matching.

Here all children DNA's are collected and taken as a dataset. The parent's DNA is taken as input and is compared with all child's DNA in the dataset. If a match is found then the child is related to particular parent.

II. PROPOSED SYSTEM

The idea of this project came into existence because of the increasing number of child missing cases. The objective of this project is to design an algorithm using machine learning by which a match between the DNA of matching child and parent can be identified.

We will use short Tandem repeats and Bayes classification algorithm to match the DNA's with these algorithm we are able to extract the repeating sequence of DNA and compare them to find the matching DNA's.

III. SYSTEM REQUIREMENTS

1. Hardware Requirements

The minimum hardware requirements to execute the system are as follows:

- Processor - Intel I3
- RAM - 4GB and more
- Storage - 1GB

2. Software Requirements

- Operating System – Windows 10
- Backend – Python
- Frontend – HTML

3. Functional Requirements

- Data Collection
- Data Preprocessing
- Training and Testing
- Modeling
- Predicting

4. Non-Functional Requirements

- Performance
- Reliability
- Availability
- Security
- Maintainability
- Portability

IV. METHODOLOGY

The procedure and methodology involved at first, we need to collect all the DNA’s of retraced children and create a dataset with these DNA’s. When a parent comes his DNA is collected. This parent DNA is compared with each and every children in the child DNA dataset. This can be done by using two algorithms namely Short Tandem Repeats and Bayes Classification Algorithm.

SHORT TANDEM REPEATS ALGORITHM:

Short Tandem Repeats (STR) also known as micro satellites are simple sequence repeats, are Short Tandemly Repeated DNA sequences that involve a repetitive unit of 1-6BP, forming series with length of upto 100 nucleotides (nt). They appear scattered more or less evenly throughout the human genome, accounting for about 3% of entire genome. Most STRs are found in the non coding regions, while only 8% located in the coding regions. On average one STR occurs per 2000BP in the human genome. The STRs are identified in both parent and child DNA and the fig.1 are compared.

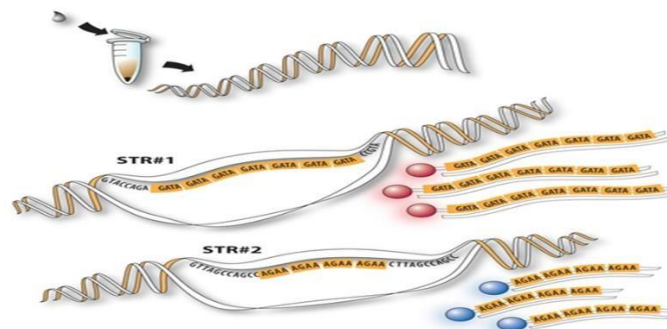


Fig 1: Short Tandem Repeats

BAYES CLASSIFICATION ALGORITHM:

Bayes classification is a supervised learning algorithm which is based on Bayes theorem and used for solving classification problems. It is a probabilistic classifier, which means it predicts on the basis of the probability.

After finding the repeated DNA sequence using Short Tandem Repeats, the probability of parents DNA with respect to child DNA is identified using Bayes Classification Algorithm.

V.IMPLEMENTATION

The method of implementation is explained by the below steps:

Step-1 Children DNA'S that are collected are taken as child DNA dataset.

Step-2 Parent's DNA report is taken as input.

Step-3 When parent's DNA is given as input, it is compared with the child DNA dataset.

Step-4 If a match is found between the parent DNA and child DNA, then the child name is displayed in the output column. If there is no match, then output is displayed as "Match Not Found".

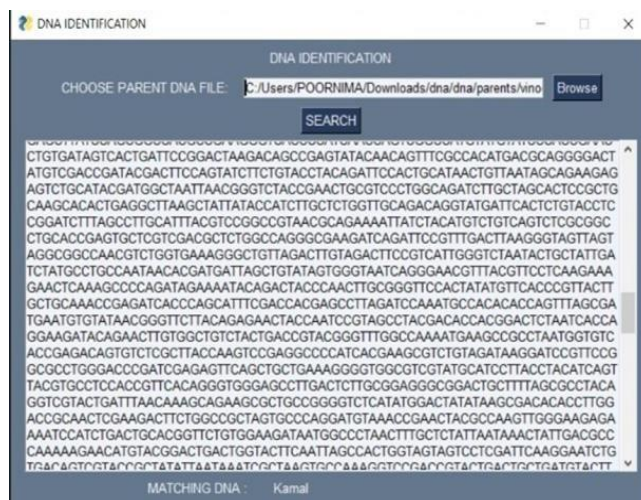


Fig 2: User Interface Screen

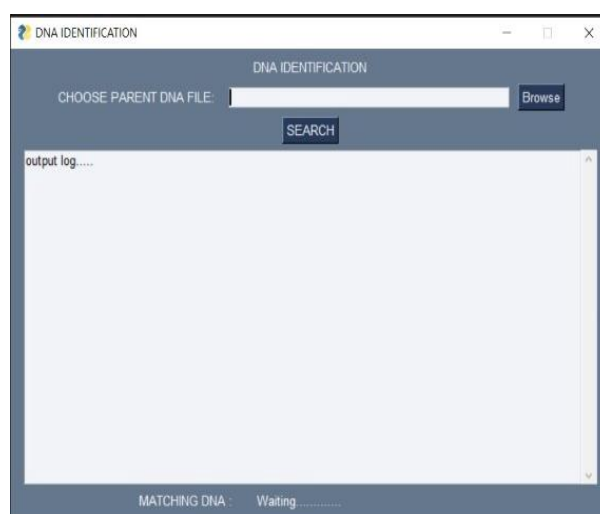


Fig 3: Input screen

VI. RESULTS

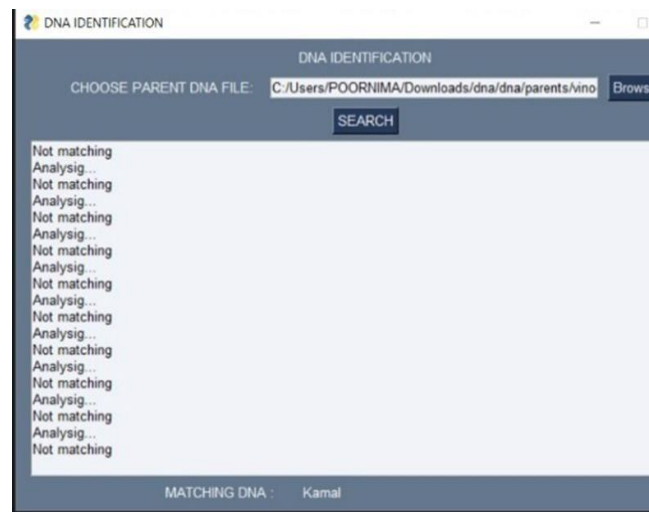


Fig 4:Output Screen



Fig 5: Output Screen

VII.CONCLUSION

This project provides the best way for identifying a child's identity in missing case. Bayes classification provides the probability of matching DNA.

Short Tandem repeats gives the repeated DNA sequences with helps in DNA matching.

REFERENCES

1. Bar, W (1994) DNA recommendations -1994 report concerning further recommendations of the DNA 160(2-3): 90-101.
2. Bar, W (1997) DNA recommendations -further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Journal of Legal Medicine 110: 175-176.
3. Benecke M (1997) DNA typing in forensic medicine and in criminal investigation's current survey 188: 181-188.
4. Bar, W (1997) DNA recommendations -further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Journal of Legal Medicine 110: 175-176.
5. Benecke M (1997) DNA typing in forensic medicine and in criminal investigation's current survey 188: 181-188.
6. Burger MF (2005) Buccal DNA samples for DNA typing: New collection and processing methods. Biotechniques 39: 257-261.
7. Burgoyne LA (1996) Solid medium and method for DNA storage. US Patent 5(496): 562.