How to Cite:
Al-Hussein, M. A. S., Al-Nima, R. R. O., & Woo, W. L. (2022). Employing deoxyribonucleic acid (DNA) for personal verification. *International Journal of Health Sciences*, *6*(S9), 126–140.
https://doi.org/10.53730/ijhs.v6nS9.12176

Employing deoxyribonucleic acid (DNA) for personal verification

Manhal Ahmad Saleh Al-Hussein
Northern Technical University, Mosul, Iraq
Corresponding author email: engmas200811@gmail.com

Raid Rafi Omar Al-Nima
Northern Technical University, Mosul, Iraq
Email: raidrafi1@gmail.com

Wai Lok Woo
Northumbria University of London, United Kingdom
Email: wailok.woo@northumbria.ac.uk

***Abstract*---**Deoxyribonucleic Acid (DNA) is one of the most important biometrics, it is especially employed for individual recognition. In this study, people are recognized based on their DNAs. An effective algorithm for acquiring the unique DNA patterns is adopted. It is called the Unique Personal DNA Pattern (UPDP). This algorithm focuses on the repeated unique pattern in each DNA sample. Four databases are utilized in this paper. These are the DNA Classification (DC), Sample DNA Sequence (SDS), Human DNA Sequences (HDS) and DNA Sequences (DS). Such interesting and remarkable results are obtained in both recognition types of verification. In verification, FARs are achieved 0.32%, 0.31%, 0% and 0.16% for the DC, SDS, HDS and DS, respectively. Whereas, all False Rejection Rates (FRRs) in verification are recorded 0% for the four databases.

***Keywords*---**DNA Biometric, Pattern, Verification, UPDP algorithm.

1 Introduction

The concept of "biometric" is currently widely used for referring to people recognition. It has been explored in various applications for many years. Examples of such applications are personal verification, criminal investigations, and security Systems (Al-Nima, 2017). Due to the increasing number of biometrics-based systems being deployed in many forensic and civilian verification applications, biometrics and their applications have attracted significant attentions. Furthermore, identifying and verifying individuals based on

their physiological or behavioural characteristics occupies a very big part in biometrics. Verification is a term that points to the agreement or disagreement of an individual identity claim after checking his/her certain information in the database, it follows a one-to-one policy (Jain *et al.*, 2006). The verification are known as recognition (ISO and IEC, 2012). Also, biometric systems that rely on human characteristics (biometrics) are much better than other systems that use personal identification numbers (PINs), passwords or codes, this is because they can easily be lost, stolen, forgotten or hacked (Jain *et al.*, 2006). Accordingly, many studies are established using different physiological or behavioural biometrics such as handwriting (Plamondon & Srihari, 2000; Anikin & Anisimova, 2016), voice features (Al-Kaltakchi *et al.*, 2020; Al-Kaltakchi *et al.*, 2020), retinal veins pattern (Zahra Waheed *et al.*, 2015; Zahra Waheed *et al.*, 2016), fingerprint (Ibrahim *et al.*, 2021; Hemalatha, 2020; PATEL *et al.*, 2021), ear print (Ali *et al.*, 2021b; Ali *et al.*, 2021; Ali *et al.*, 2021a), face appearance (Al-Nima *et al.*, 2019; Mustafa & Al-nima, 2009), iris print (Al-Nima, 2006; Khalil *et al.*, 2009), palm print (Al-Nima *et al.*, 2019; Albak *et al.*, 2021) and finger texture (Al-Nima *et al.*, 2020; Al-Kaltakchi *et al.*, 2019).

Deoxyribonucleic Acid (DNA) is one of the most common physiological biometrics that holds unique code for each individual. It provides reliable information about personal identity. DNA provides accurate and highly efficient personal recognition, its data does not change during an individual's life or even after death. Therefore, it is used in the field of forensic science (Hashiyada, 2011). It has all the properties that make it perfectly been considered as a biometric (i.e. uniqueness, continuity and universality (Abaza & Ross, 2010)).

There are about 60 trillion cells in a human's body (Hashiyada, 2011). Each cell contains the same DNA genetic information (Butler, 2005). In addition, the DNA has essential information for building the body parts (Hashiyada, 2011).

The DNA contains a big quantity of human genomes. Distinct regions have been determined in a DNA, they are polymorphous and differ from one person to another. Furthermore, the precise region of polymorphic DNA to analyse varies depending on the instance and the technology available. In every personal cell, two complementary DNA genomes are located within its nucleus (Goodwin *et al.*, 2007). All necessary data for a human structure is contained in the DNA. Also, DNA data offers useful information for important fields such as environmental sciences, medical sciences, forensics and historical researches (Jain *et al.*, 2006).

It has been discovered that certain regions in the DNA contain unique repeated patterns (sequences). Thus, the number of repeated patterns in a single DNA sample differs from one individual to another (Butler, 2005).

Identical twins (or monozygotic twins) completely and accurately share genomes, so, they cannot be distinguished by the DNA polymorphism. However, one identical twin may sometimes have developed schizophrenia or cancer and the other may have not (Zwijnenburg *et al.*, 2010). Moreover, environmental factors can alter the gene and its vulnerability to have the disease by affecting its compositions. So, it has recently been revealed that identical twins have significant differences in their DNA sequences pattern (Haque et al., 2009).

The DNA is actually two strands twisted together in the format of a double helix due to the phenomenon of “base pairing” (Butler, 2005). The DNA strand is a polymer of nucleotides, where each nucleotide consists of three parts: sugar, phosphate and base (Hashiyada, 2011). Within the DNA structure, information is encoded on the sugar and phosphate by a sequence of four different nitrogenous bases: Adenine (A), Guanine (G), Thymine (T) and Cytosine (C) (Goodwin *et al.*, 2007). The four symbols (A, G, T and C) represent the content in each nucleotide. Therefore, the difference in the sequence of the nucleotide indicates biodiversity between all living creatures and not only between humans (Hashiyada, 2011).

For each DNA loci, the (A, T, C or G) are four possibilities. They all can literally provide trillions of combinations. Also, based on the “base pairing” rule between DNA strands, a nitrogenous base in a nucleotide unit of one strand is linked to its corresponding “complementary base” in a nucleotide unit of the other strand (Butler, 2005). By the means of hydrogen bonds, a single nucleotide in the any strand is connected to a nucleotide in the other strand (the complementary base) (Goodwin *et al.*, 2007). The hydrogen bonds are of two types, where the base C is linked to the base G through three hydrogen bonds and the base A is linked to the base T through two hydrogen bonds. So, the bonds between the bases C and G are slightly stronger than the bonds between the bases A and T. Also, The two strands of DNA are "anti-parallel" directions, which means that if one of them is in the 3' to 5' direction, then the other will be in the 5' to 3' direction. In other words, the two strands are in opposite in directions. Therefore, for the sequence of a single DNA strand, the complementary sequence can simply be expected based on the basic pairing rules of: A and T are always connected together, are G and C are always complemented to each other (Butler, 2005), see Figures 1 (Goodwin *et al.*, 2007) and 2 (Butler, 2005).

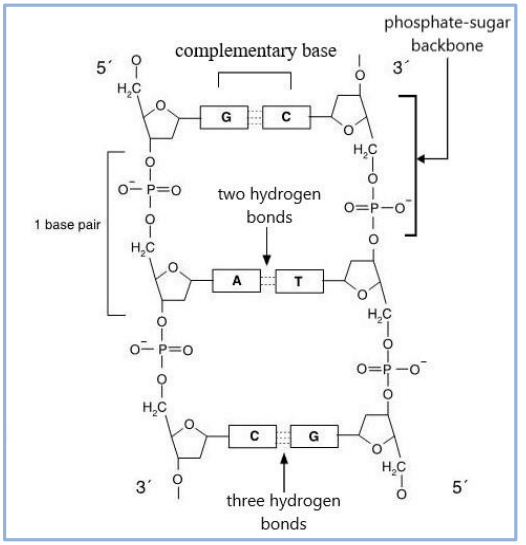


Figure 1. DNA chemical structure, where it shows the molecular structure of two DNA strands (Goodwin *et al.*, 2007)

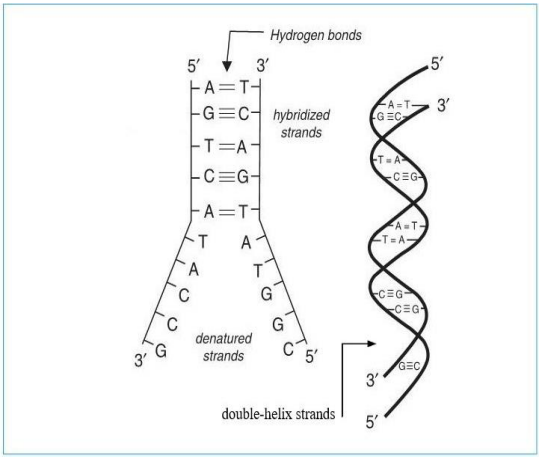


Figure 2. Basic DNA structure, where the left part demonstrates nitrogenous bases and hydrogen bonds, and the right part shows the two strands of a DNA with their helical twisting (Butler, 2005)

The aim of this work is to apply the DNA for identifying people. The contribution here is represented by providing an effective algorithm for acquiring the unique DNA patterns of the A, T, C and G. This proposed algorithm is called the Unique Personal DNA Pattern (UPDP). The remaining sections of the paper are organized as follows: a literature review is presented in Section 2, the UPDP theory is described in Section 3, experimental results are discussed in Section 4 and the conclusion is provided in Section 5.

Literature Review

For the literature, there are many studies that focused on the DNA. Introduced a new DNA-based algorithm that focused on the iris of eye. This paper was conducted at the Madrid-Spain. It specifically concentrated on the identification and documentation of irises. It gave a new approach for iris detection by aligning two iris models in a biometrics system. In the used new, approach the DNA alignment, then, fuzzy logic to apply the identification comparisons. This could significantly reduce the identification time and also provided the ability to recognize a specific user under low-quality conditions such as images of blurred eyes (Sierra *et al.*, 2008).

Approached a new DNA sequencing compressor based on the pattern recognition. This method was implemented in India. All genetic information are encoded in a DNA for all living subjects. An increasing number of DNA sequences were discovered and stored in genetic databases. The sizes of databases were expected to grow at an exponential rate. Compression was desirable to reduce storage requirements as well as transmission time. The pattern recognition-based DNA sequencing compression algorithm was proposed. It compressed DNA sequences by detecting related patterns to have a greater compression ratio and good compression gain than many conventional compressors, such as RARLabs' WinRAR version 4.x. The proposed approach would be mostly useful for scientists who work with genetic data (Arokiaraj & Robert, 2012).

Focused on the DNA methylation in the case of discordant breast cancer. DNA methylation is a biological process in which methyl groups are added to a DNA molecule. The process of methylation can change the activity of a segment of DNA without changing its primary sequence and DNA methylation usually prevents the transcription of genes. This work was perform on DNA samples from patients with breast tumors, the participants were all from Caucasian in origin. In particular, the recognition of identical twins was emphasized for a docking protein. DNA hyper-methylation was determined to be consistently detectable in three independent sample settings: blood taken from breast cancer incompatible twins, primary breast tumors and breast cancer cell lines (Heyn *et al.*, 2013).

Determined the DNA sequence by using stream matching techniques. This paper was established at the Teerthanker Mahaveer University (TMU) in Moradabad, India. Frameworks were developed for the DNA sequencing in the field of bioinformatics which has become the subject of many research studies. Two algorithms were utilized and discussed, these are: the Rabin-Karp (R-K) algorithm and the Maximum Common Sub-stream (MCS) algorithm (Tripathi & Pandey, 2016).

Designed a DNA-based biosecurity system. This work was performed at Ladoka Akintola University of Technology (LAUT) in Obumuso, Nigeria. In this study, the DNA accomplished for examination conduct. The primary method of a biometric system was to gather biometric data via sensor components. After extracting biometric data by the module, comparisons for identifying subjects were applied. The value of Variable Number Tandem Repeat (VNTR), which repeats at a number of unique loci, was revealed by the DNA profiling. Templates were encrypted utilizing algorithmic alteration of biometric materials. Then the STR value was determined resulting in a unique personal identity information that would be utilized for statistical and theoretical analysis in the final recognition (Afolabi & Akintaro, 2017).

Suggested a method for situ DNA replication and detection in microchannels from Laser-induced heating. This method was performed in Taiwan. This work was suggested through laser-induced heating and high avidin–biotin binding, a method for in situ local DNA replication and detection in a long DNA strand. Di electrophoresis was used to stretch and immobilize DNA strands on both ends of the electrode in order to accomplish the desired DNA replication. Following that, thermal cycles caused by laser-induced heating to repeat local DNA sequences. The duplicated local DNA sequence could be recognized after six laser-induced heat cycles. The proposed method should make biosample gene sequence analysis more efficient (Hung & Chen, 2018).

Introduced a method for compressing DNA sequencing by the application of signal processing. This paper was conducted at the Hong Kong Polytechnic University (HKPU) in Hong Kong. Properties of DNA sequencing that could be utilized for compression were explored. Two stress theories were used: reference-based methods and reference-free methods. In reference-based methods, repeating patterns were determined between reference sequences and the DNA sequence target. In reference-free approaches, redundancies within the specific DNA sequence to be compressed were investigated. The study was mainly focused on

the population sequences for each category generated by the Next Generation Sequencing (NGS) (Law, 2019).

Provided a scalable algorithm to interpret DNA sequences and predict killer T-cell responses in Systemic Lupus Erythematosus (SLE) patients. This study was established in Nigeria. The purpose of this study was to develop an algorithm that could assess a patient's DNA sequence and predict killer T-cell responses for people with the SLE. The procedure was to find a gene variant in a DNA patient. For the matching, an approximation matching algorithm based on the Boyer-Moore (BM) algorithm was utilized. The method could anticipate killer T-cell responses, allowing better early detection and treating SLE patients (Ephraim *et al.*, 2022).

From the literature, it can be explored that there are no sufficient DNA studies that consider Iraq as the case study, although this country is significantly influenced by wars as too many people die and lost. In addition of increasing corruptions and criminals in Iraq during the last decades. So, this paper is provided to address a vital and very sensitive DNA topic of recognition by doing such work in this country, which may later benefit from this study.

2 Materials and Methods

2.1 Descriptions

One of the most important biometrics that has spread during the last decade for revealing human's identity is the DNA. According to relevant scientific studies and researches, a single personal cell contains 23 pairs of chromosomes within its nucleus. Each chromosome contains a large number of genes. One gene contains two strands of the DNA. The unit measurement for the DNA length is a base pair (bp) and the length of a human's DNA is approximately 3.2G bp. Any living creature consists of certain regions as given in Figure 3 (Goodwin *et al.*, 2007).

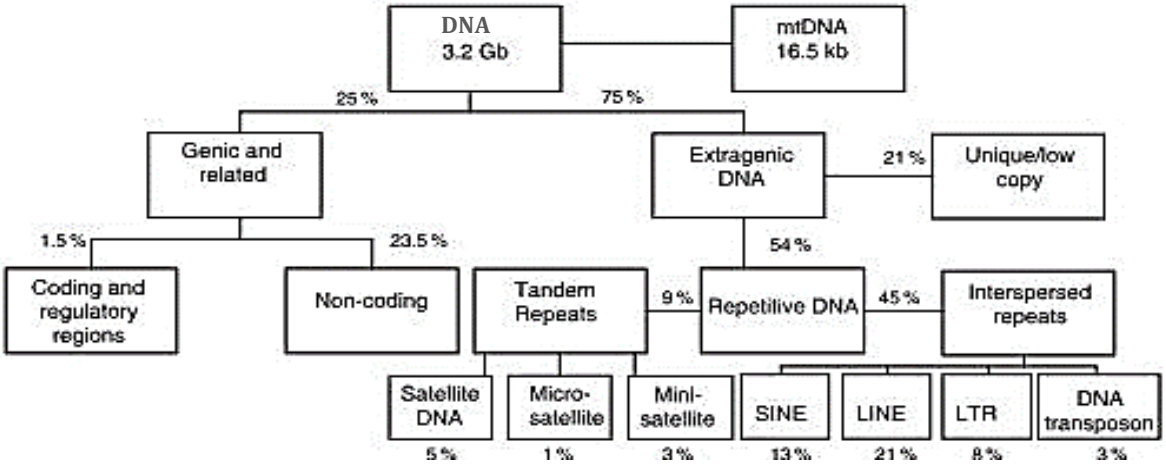


Figure 3. Determined regions of the DNA as shown in (Goodwin *et al.*, 2007)

Human's DNA, which is located inside the cell nucleus, is divided into several parts. The first part is the mitochondrial DNA (mDNA), its length is 16.5 kbp. This length is considered to be small in relation to the full DNA length. Other parts are genic and related (represents 25% of full DNA length), and extragenic DNA (represents 75% of full DNA length). Genic and related DNA part is also divided into two parts. The first part is coding and regulator regions (represents 1.5% of full DNA length), and the second part is non-coding (represents 23.5% of full DNA length). Extragenic DNA is also divided into two parts. The first part is unique/low copy (represents 21% of full DNA length) and the second part is repeitive DNA (represents 54% of full DNA length). The repeitive DNA is divided into two parts. The first part is the tandem repeats (represents 9% of full DNA length) and the second part is interspersed repeats (represents 45% of full DNA length). Tandem repeats is divided into three regions, these are the DNA satellites (represents 5% of full DNA length), minisatellites (represents 3% of full DNA length) and microsatellite (represents 1% of full DNA length). Interspersed repeats is divided into four parts, these are the Short Inter-Spersed Element (SINE) (represents 13% of full DNA length), Long Interspersed Element (LINE), (represents 21% of full DNA length), Long Terminal Repeats (LTR) (represents 8% of full DNA length) and DNA transposon (represents 3% of full DNA length). Coding region responsibles for the formation of a body and its organs, it does not change during generations. Remaining DNA regions differs from one person to another without affecting the essential compositions of a body (Goodwin *et al.*, 2007).

In this paper, we will deal with the STR data which are contained in the microsatellite region. They are small in size and consist of groups of nucleotides which are repeated side by side, and they called alleles. Core of repeated pattern (sequenced symbols of nucleotides) in an allele can be of the size from 1 to 6 bp. The name of allele is given based on its contained repeated patterns, see Figure 4 (Goodwin *et al.*, 2007).

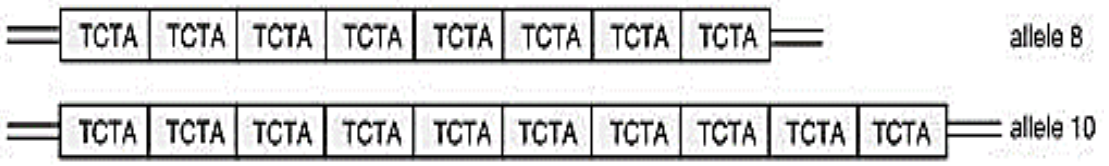


Figure 4. Examples of STRs, which indicates two alleles (Goodwin *et al.*, 2007)

2.2 UPDP Algorithm

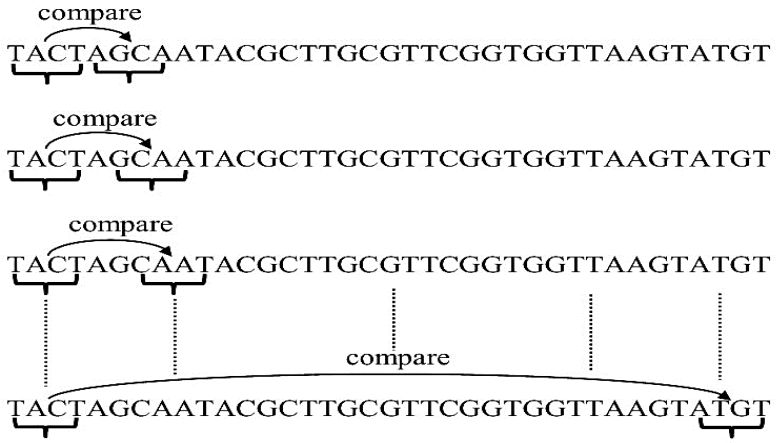
A proposed algorithm called the UPDP is proposed. Its structure starts with samples of DNA data that are acquired by biosensors. They have the form of sequenced symbols of nucleotides. Each DNA sequence is considered as an input and is treated as form of characters. Figure 5 demonstrates a Sample of a DNA sequence, it consists of symbols of nucleotides (Goodwin *et al.*, 2007).

AGCTGTAAGTCTATACGTATCGTTAGTGCCTTGACTATGTCCGTA

Figure 5. Sample of a DNA sequence, it consists of symbols of nucleotides (Goodwin *et al.*, 2007)

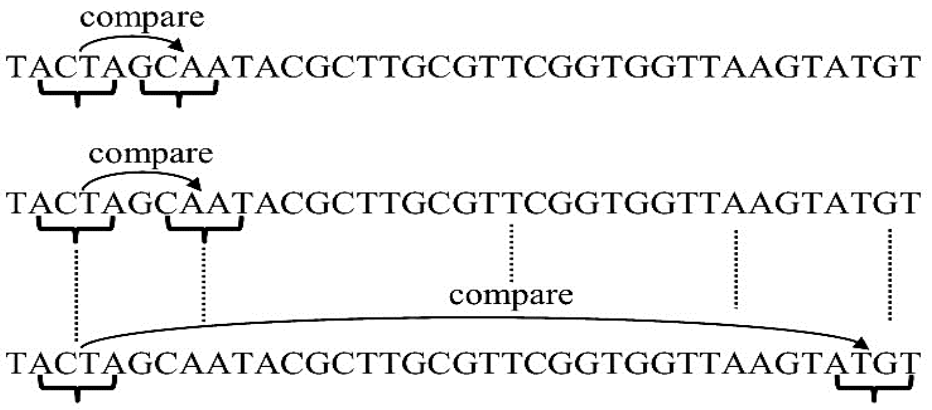
As mentioned, the symbols A, C, T and G represent the four nucleotides in a DNA. In this work, the quadruple STR named the allele is employed. It is important and common for counting the number of repeated nucleotides pattern (Asogawa *et al.*, 2020). UPDP algorithm considers applying many comparisons for extracting the repeated nucleotides pattern in a DNA sequence. The algorithm is demonstrated with the following example:

Step 1: Applying comparisons for the first quadruple pattern in a DNA sequence or DNA sample with counting the number of its frequency:



Step 2: Saving the number of frequenting the first quadruple pattern in the DNA sequence.

Step 3: Shifting the comparisons for the second quadruple pattern in a DNA sequence with counting the number of its frequency:



Step 4: Saving the number of frequenting the second quadruple pattern in the DNA sequence.

Step 5: Repeating step 3 to step 4 for all other quadruple patterns in the DNA sequence with counting the numbers of their frequencies. The final comparison can be demonstrated as follows:

Step 6: The numbers of frequenting all quadruple patterns in the DNA sequence c_i is considered for all probabilities $P(A)$ of the quadruple patterns A. Then, the following equation is utilized:

$$C = \max(c_i) , \quad i=1,2,\ldots,m \qquad (1)$$

Where C is the maximum number of frequenting all quadruple patterns c_i, max is the maximum operation and m is the number of quadruple patterns probabilities, which is equal to 256.

Step 7: Repeating steps 1 to 6 for other DNA sequences or DNA samples in the employed template of a database. Saving the maximum numbers of frequenting all quadruple patterns C_j for all provided DNA sequences, where $j= 1,2,\ldots, n$ and n is the number of DNA samples in the employed template of a database.

Step 8: Establishing a recognition code by exploiting the following equation:

$$D_j = B_j \times 100 + C_j \qquad (2)$$

where D_j is the established recognition code and B_j is the index number in the probability of $P(A)$.

Step 9: Applying the verification process by considering the obtained establishing codes.

Figure 6 shows the general UPDP algorithm block-diagram.

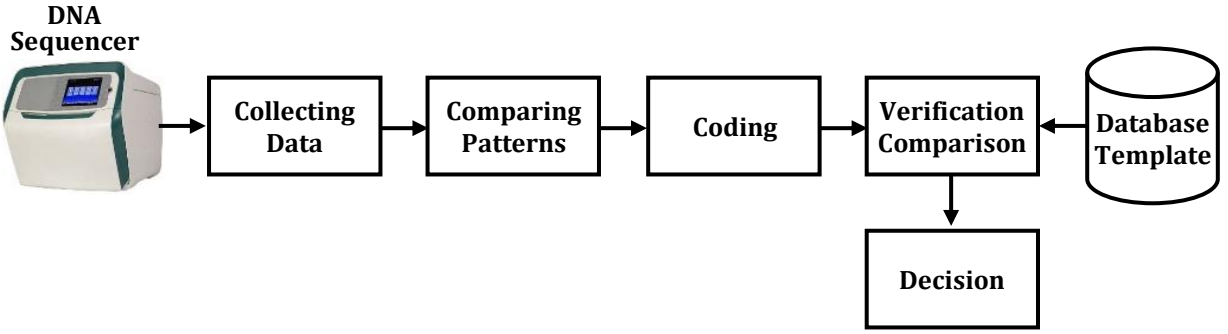


Figure 6. The general UPDP algorithm block-diagram

2.3 Verification

Each one of the obtained codes from the UPDP algorithm is considered as an assigned personal identity number. Hereafter, it is possible to perform the verification operation.

The practice of linking a specific individual with an identity is known as personal identification. In other words, it is a process of finding the identity of a particular individual in a whole database (answer to a question "Who am I?"). Verification requires verifying the claimed of identity (answer to a question "Am I who I claim to be?") (Jain *et al.*, 2000).

For the verification process, a comparison is made to verify the validity of the claimant's claim by matching the identity of the claimant with a certain stored identity in the database template (Al-Nima, 2017). It follows a one-to-one policy (Jain *et al.*, 2006). Standard Euclidean Distance (SED) is used to calculate the distance between the provided identity code and stored identity code in the database template. The performance of the proposed UPDP algorithm is determined according to the SED values by considering metrics of the False Accept Rate (FAR) and False Reject Rate (FRR). They are two widely used standard indicators for measuring biometric performances. The FAR is the ratio of imposters who are wrongly allowed acceptance. The FRR is the ratio of valid users who are wrongly denied acceptance (Ko, 2005). FAR and FRR equations can respectively be represented as follows (Goh *et al.*, 2010):

$$FAR = \frac{Number\ of\ Accepted\ Imposters}{Total\ Number\ of\ Imposters} \times 100\% \quad (4)$$

$$FRR = \frac{Number\ of\ Rejected\ Clients}{Total\ Number\ of\ Clients} \times 100\% \quad (3)$$

3 Results and Discussions

Four databases were employed in this study. Firstly, a database from ("DNA-Classification," n.d.) called the DNA Classification (DC). Secondly, a database from ("Sample DNA Sequence," n.d.) named the Sample DNA Sequence (SDS). Thirdly, a database from ("Human DNA Sequences," n.d.) called the Human DNA Sequences (HDS). Fourthly, a database from ("DNA Sequence," n.d.) named the DNA Sequences (DS). Each one of the three databases consists of sequences of the DNA symbols (A, G, T and C). Big numbers of samples are provided in the four databases. That is, the DC compromises of 106 samples, the SDS composes of 426 samples, the HDS has many more samples, however, only 500 samples are used in this paper and the DS also consists of many more samples, but only 1000 samples are used in this study.

Verification is considered here for all of the four employed databases. The used machine for all experiments has the following specifications: laptop computer of type HP, Central Processing Unit (CPU) of type Intel (R) Core (TM) i5-5200U, CPU speed of 2.20 GHz, Random Access Memory (RAM) of size 8 GB, Double Data Rate (DDR) RAM of version 3, internal graphics card of type Intel (R) HD Graphic 5500 and graphics memory capacity of size 1 GB.

Big numbers of comparisons have been evaluated in the case of verification. To illustrate, the number of clients for the DC is equal to 106 samples, whereas, the number of imposters is reached to 309 samples. Similarly, the number of clients for the SDS is equal to 426 samples, whilst, the number of imposters is reached to 1264 samples. Likewise, the number of clients for the HDS is equal to 500 samples, however, the number of imposters is reached to 1498 samples.

Correspondingly, the number of clients for the DS is equal to 1000 samples, nevertheless, the number of imposters is reached to 2974 samples. Table 1 shows verification performances for the various employed databases by using the suggested UPDP approach.

Table 1
Verification performances for the various employed databases by using the suggested UPDP approach

Database	Number of Clients (samples)	Number of Imposters (samples)	FAR (%)	FRR (%)
DC	106	309	0.32	0
SDS	426	1264	0.31	0
HDS	500	1498	0	0
DS	1000	2974	0.16	0

This table shows that the DC database with the number of clients 106 samples and number of imposters 309 samples attains the performances of FAR = 0.32% and FRR = 0%. The SDS database with the number of clients 426 samples and number of imposters is 1264 samples achieves the errors of FAR = 0.31% and FRR = 0%. The HDS database with the number of clients 500 samples and number of imposters 1498 samples obtains the percentages of FAR = 0% and FRR = 0%. The DS database with the number of clients 1000 samples and number of imposters 2974 samples benchmarks the results of FAR = 0.16% and FRR = 0%.

To sum up, the performances of employed verifications show successful and acceptable error percentages for the four databases by applying the proposed UPDP approach. The best result is benchmarked for the HDS database in verification where it recorded FAR=0% and FRR=0%, and obviously this is a remarkable performance. It is also worth highlighting that the FRR has attained zero percentages for all of our experiments in verifications and for all of the four databases.

4 Conclusion

In this paper, we proposed and presented the UPDP algorithm approach for recognizing individuals based on their DNA samples. Since each person has a unique and distinct number of repeated pattern in a DNA sample, the UPDP approach is suggested for providing a personal identity to be used for verification comparison. Four databases from the DC, SDS, HDS and DS were employed. Interesting and remarkable results were obtained as the FRRs were recorded as zero percentages for all the employed databases in the verification. The FARs achieved 0.32%, 0.31%, 0% and 0.16% for the DC, SDS, HDS and DS in verification processes. The best result was benchmarked for the HDS database in verification, where it obtained FAR = 0% and FRR = 0%. This is obviously an outstanding performance. It is also worth mentioning that the FRR achieved zero percentages in all verifications for all of the four employed databases. It can be yield that the proposed UPDP algorithm is successful and effective.

Acknowledgments
We appreciate that the Ministry of Health /Forensic Medicine Directorate provided us with the information we needed to complete this study.

References

A. O. Afolabi, & O. M. Akintaro. (2017). Design of DNA Based Biometric Security System for Examination Conduct. *Journal of Advances in Mathematics and Computer Science*, *23*(6), 1–7. https://doi.org/10.9734/JAMCS/2017/27251.

Abdulsattar M. Ibrahim, Abdulrahman K. Eesee, & Raid Rafi Omar Al-Nima. (2021). Deep fingerprint classification network. *Telkomnika (Telecommunication Computing Electronics and Control)*, *19*(3), 893–901. https://doi.org/10.12928/TELKOMNIKA.v19i3.18771.

Anil Jain, Lin Hong, & Sharath Pankanti. (2000). BIOMETRIC IDENTIFICATION. *Communications of the Association for Computing Machinery*, *43*(2), 91–98.

Anil K. Jain, Ruud Bolle, & Sharath Pankanti. (2006). BIOMETRICS: Personal Identification in Networked Society. In *Biometrics* (pp. 213–229). Springer US. https://doi.org/10.1007/0-306-47044-6_10.

ASHA K. PATEL, UNNATI P. PATEL, & Falguni Suthar. (2021). Fingerprint Recognition in Biometric Security Systems. *JOURNAL OF INFORMATION, KNOWLEDGE AND RESEARCH IN COMPUTER SCIENCE AND APPLICATIONS FINGERPRINT*, *01*(02), 44–49. https://doi.org/10.13140/RG.2.2.11683.17441.

Ayman Abaza, & Arun Ross. (2010). Towards understanding the symmetry of human ears: A biometric perspective. *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 1–7. https://doi.org/10.1109/BTAS.2010.5634535.

Bonnie Ngai-Fong Law. (2019). Application of signal processing for DNA sequence compression. *IET Signal Processing*, *13*(6), 569–580. https://doi.org/10.1049/iet-spr.2018.5392.

DNA-classification. (n.d.). In *[on line]*. https://www.kaggle.com/code/ashishpawar511/dna-classification/ data.

DNA Sequence. (n.d.). In *[on line]*. https://www.kaggle.com/fsiamp/dna-sequences?select=1158776424248. data.

F. Nipa Haque, Irving I. Gottesman, & Albert H.C. Wong. (2009). Not really identical: Epigenetic differences in monozygotic twins and implications for twin studies in psychiatry. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, *151C*(2), 136–141. https://doi.org/10.1002/ajmg.c.30206.

Hemalatha S. (2020). A systematic review on Fingerprint based Biometric Authentication System. *2020 International Conference on Emerging Trends in Information Technology and Engineering (Ic-ETITE)*, 1–4. https://doi.org/10.1109/ic-ETITE47903.2020.342.

Heyn, H., Carmona, F. J., Gomez, A., Ferreira, H. J., Bell, J. T., Sayols, S., Ward, K., Stefansson, O. A., Moran, S., Sandoval, J., Eyfjord, J. E., Spector, T. D., & Esteller, M. (2013). DNA methylation profiling in breast cancer discordant identical twins identifies DOK7 as novel epigenetic biomarker. *Carcinogenesis*, *34*(1), 102–108. https://doi.org/10.1093/carcin/bgs321.

Human DNA Sequences. (n.d.). In *[on line]*. https://www.kaggle.com/datasets/sooryaprakash12/human-dna-sequences?select=human_data.txt.

Hung, M., & Chen, C. (2018). Laser-induced heating for in situ DNA replication

and detection in microchannels. *IET Nanobiotechnology*, *12*(6), 841–845. https://doi.org/10.1049/iet-nbt.2017.0302.

Igor V. Anikin, & Ellina S. Anisimova. (2016). Handwritten signature recognition method based on fuzzy logic. *2016 Dynamics of Systems, Mechanisms and Machines (Dynamics)*, 1–5. https://doi.org/10.1109/Dynamics.2016.7818968.

ISO and IEC. (2012). Information technology vocabulary part 37: Biometrics. *International Standard ISO/IEC 2382-37*, 1–21.

John M. Butler. (2005). Forensic DNA Typing. In *Elsevier Academic Press; 2nd edition (March 8, 2005)*. Elsevier Academic Press. https://www.elsevier.com/books/forensic-dna-typing/butler/978-0-08-047061-0.

Lubab H. Albak, Raid Rafi Omar Al-Nima, & Arwa Hamid Salih. (2021). Palm print verification based deep learning. *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, *19*(3), 851–857.

Masaki Hashiyada. (2011). DNA biometrics. In Yang Dr. Jucheng (Ed.), *Biometrics* (pp. 1–266). InTech. https://doi.org/10.5772/18139.

Mazin R. Khalil, Mahmood S. Majeed, & Raid R. Omar. (2009). Personal Identification with Iris Patterns. *AL-Rafidain Journal of Computer Sciences and Mathematics, College of Computer Sciences and Math / University of Mosul / Iraq*, *6*(1), 13–26.

Masruroh, M., Soetrisno, S., Wijaya, M., & Anantanyu, S. (2021). Analysis of stakeholder's role in community in the effort of early-age marriage prevention: a case study in junior high school students in Gunung Kidul District Yogyakarta. International Journal of Health & Medical Sciences, 4(3), 298-303. https://doi.org/10.31295/ijhms.v4n3.1749

Michael K.O. Goh, Connie Tee, & Andrew B.J. Teoh. (2010). Bi-Modal Palm Print And Knuckle Print Recognition System. *Journal of IT in Asia*, *3*(1), 53–66. https://doi.org/10.33736/jita.37.2010.

Minoru Asogawa, Ayumu Ohno, So Nakagawa, Eriko Ochiai, Yasuhiro Katahira, Megumi Sudo, Motoki Osawa, Masatoshi Sugisawa, & Tadashi Imanishi. (2020). Human short tandem repeat identification using a nanopore-based DNA sequencer: a pilot study. *Journal of Human Genetics*, *65*(1), 21–24. https://doi.org/10.1038/s10038-019-0688-z.

Musab T. Al-Kaltakchi, Raid R.Omar, Hikmat N. Abdullah, Tingting Han, & Jonathon A. Chambers. (2019). Finger Texture Verification Systems Based on multiple spectrum Lighting Sensors with Four Fusion Levels. *Iraqi Journal of Information & Communications Technology*, *1*(3), 1–16. https://doi.org/10.31987/ijict.1.3.28.

Musab T. S. Al-Kaltakchi, Raid Rafi Omar Al-Nima, Mahmood Alfathe, & Mohammed A. M. Abdullah. (2020). Speaker Verification Using Cosine Distance Scoring with i-vector Approach. *2020 International Conference on Computer Science and Software Engineering (CSASE), IEEE*, 157–161. https://doi.org/10.1109/CSASE48920.2020.9142088.

Musab T. S. Al-Kaltakchi, Raid Rafi Omar Al-Nima, & Mohammed A. M. Abdullah. (2020). Comparisons of extreme learning machine and backpropagation-based i-vector approach for speaker identification. *TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES*, *28*(3), 1236–1245. https://doi.org/10.3906/elk-1906-118.

Mustafa, F. S., & Al-nima, R. R. (2009). Face Recognition Using Invariant Moments Features. *Tikrit Journal of Pure Science*, *14*(2), 253–259.

Nwoye O Ephraim, Fidelis P Obinna, Nwosu O I, Balogun O Jessy, Raid Rafi Al-Nima, & Wai Lok Woo. (2022). A Scalable Algorithm for Interpreting DNA Sequence and Predicting the Response of Killer T-Cells in Systemic Lupus Erythematosus Patients. *The International Journal of Bioinformatics and Intelligent Computing (IJBIC)*, *1*(1), 43–65.

Panneer Arokiaraj, S., & Robert, L. (2012). Pattern Recognition based DNA Sequence Compressor. *2012 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2012*. https://doi.org/10.1109/ICCIC.2012.6510211.

Petra J.G. Zwijnenburg, Hanne Meijers-Heijboer, & Dorret I. Boomsma. (2010). Identical but not the same: The value of discordant monozygotic twins in genetic research. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *153*(6), 1134–1149. https://doi.org/10.1002/ajmg.b.31091.

Raid R. O. Al-Nima, Farqad H. Abdulraheem, & Moatasem Y. Al-Ridha. (2019). Using Hand-Dorsal Images to Reproduce Face Images by Applying Back propagation and Cascade-Forward Neural Networks. *2019 2nd International Conference on Electrical, Communication, Computer, Power and Control Engineering (ICECCPCE)*, 97–102.

Raid R. O. Al-Nima, Tingting Han, Taolue Chen, Satnam Dlay, & Jonathon Chambers. (2020). *Finger Texture Biometric Characteristic: a Survey*. 1–17.

Raid Rafi Omar Al-Nima. (2006). *Personal Identification Based on Iris Patterns*. Master thesis, Technical Engineering College of Mosul, Iraq.

Raid Rafi Omar Al-Nima. (2017). *Signal Processing and Machine Learning Techniques for Human Verification Based on Finger Textures* (Issue November). PhD thesis, Newcastle University, UK.

Raid Rafi Omar Al-Nima, Moatasem Yaseen Al-Ridha, & Farqad Hamid Abdulraheem. (2019). Regenerating face images from multi-spectral palm images using multiple fusion methods. *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, *17*(6), 3110. https://doi.org/10.12928/telkomnika.v17i6.12857.

Rejean Plamondon, & Sargur N. Srihari. (2000). Online and off-line handwriting recognition: a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 63–84. https://doi.org/10.1109/34.824821.

Sample DNA Sequence. (n.d.). In *[on line]*. https://www.kaggle.com/sreshta140/covid19-genome-sequence?select=sequence.fasta.

Sarah O. Ali, Raid R. O. Al-Nima, & Emad A. Mohammed. (2021). Individual Recognition with Deep Earprint Learning. *2021 International Conference on Communication & Information Technology (ICICT), IEEE*, 304–309. https://doi.org/10.1109/ICICT52195.2021.9568410.

Sarah Othman Ali, Raid Rafi Omar Al-Nima, & Emad Ahmed Mohammed. (2021a). Communication Establishment Based on Authenticating Earprints. *International Journal of Future Generation Communication and Networking*, *14*(1), 3242–3264.

Sarah Othman Ali, Raid Rafi Omar Al-Nima, & Emad Ahmed Mohammed. (2021b). Earprint Authentication for Communicating Purpose. *LAP Lambert Academic Publishing*, 73183.

Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2022). Post-pandemic health and its sustainability: Educational situation. International Journal of Health

Sciences, 6(1), i-v. https://doi.org/10.53730/ijhs.v6n1.5949

Alberto de Santos Sierra, Carmen Sánchez Ávila, & Vicente Jara Vera. (2008). A fuzzy DNA-based algorithm for identification and authentication in an Iris Detection System. *2008 42nd Annual IEEE International Carnahan Conference on Security Technology*, 226–232. https://doi.org/10.1109/CCST.2008.4751306.

Teddy Ko. (2005). Multimodal Biometric Identification for Large User Population Using Fingerprint, Face and Iris Recognition. *34th Applied Imagery and Pattern Recognition Workshop (AIPR'05)*, 218–223. https://doi.org/10.1109/AIPR.2005.35.

Tripathi, S., & Pandey, A. K. (2016). Identifying DNA sequence by using stream matching techniques. *2016 International Conference System Modeling & Advancement in Research Trends (SMART)*, 334–338. https://doi.org/10.1109/SYSMART.2016.7894545.

William Goodwin, Adrian Linacre, & Sibte Hadi. (2007). An introduction to forensic Genetics. In *General dentistry* (Vol. 61, Issue 5). John Wiley & Sons, Ltd 1st.ed. https://www.wiley.com/en-us/An+Introduction+to+ Forensic+Genetics-p-9780470010273.

Zahra Waheed, Amna Waheed, & M. Usman Akram. (2016). A robust non-vascular retina recognition system using structural features of retinal image. *2016 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, 101–105. https://doi.org/10.1109/IBCAST.2016.7429862.

Zahra Waheed, M. Usman Akram, Amna Waheed, & Arslan Shaukat. (2015). Robust extraction of blood vessels for retinal recognition. *2015 Second International Conference on Information Security and Cyber Forensics (InfoSec)*, 1–4. https://doi.org/10.1109/InfoSec.2015.7435497.