INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2024, VOL. 70, NO. 2, PP. 481–487
Manuscript received October 3, 2023; revised June, 2024. doi: 10.24425/ijet.2024.149569

DNA recognition using Novel Deep Learning Model

Musab T.S. Al-Kaltakchi, Hasan A. Abdulla, and Raid Rafi Omar Al-Nima

***Abstract*—DNA, a significant physiological biometric, is present in all human cells like hair, blood, and skin. This research introduces a new approach called the Deep DNA Learning Network (DDLN) for person identification based on their DNA. This novel Machine Learning model is designed to gather DNA chromosomes from an individual's parents. The model's flexibility allows it to expand or contract and has the capability to determine one or both parents of an individual using the provided chromosomes. Notably, the DDLN model offers quick training in comparison to traditional deep learning methods. The study employs two real datasets from Iraq: the Real Iraqi Dataset for Kurds (RIDK) and the Real Iraqi Dataset for Arabs (RIDA). The outcomes demonstrate that the proposed DDLN model achieves an Equal Error Rate (EER) of 0 for both datasets, indicating highly accurate performance.**

***Keywords*—DNA Recognition; Deep Learning; DNA Identification.**

I. Introduction

DNA biometrics, often known as DNA fingerprinting or DNA profiling, involved utilizing an individual's unique DNA characteristics for identification and authentication. The method relied on the distinct genetic information present in an individual's DNA to establish their identity, similar to how traditional biometric methods like fingerprinting, iris scanning, and face recognition relied on distinguishing physical attributes. Specific nucleotide sequences, including adenine, guanine, cytosine, and thymine, were present in a person's DNA, collectively constituting their genetic code [1]–[3].

Humans shared most of their DNA sequences; however, some parts of the genome showed variation across people. DNA profiles unique to individuals were constructed using these variable regions known as polymorphisms. The process of DNA biometrics involved collecting a DNA sample, often from blood, saliva, or hair, and analyzing specific DNA regions with significant variation among individuals. Gel electrophoresis and Polymerase Chain Reaction (PCR) were typically employed to analyze these regions. The resulting DNA profile served as an individual's distinct identity, often represented as a set of bands or alleles. DNA biometrics found wide-ranging applications, including paternity testing, identifying missing persons, monitoring animal populations, and forensic science, where it helped identify suspects using biological evidence recovered from crime scenes. Given the highly personal and sensitive nature of genetic information, it's essential to emphasize that the use of DNA biometrics requires careful management of privacy and ethical concerns [4]–[7].

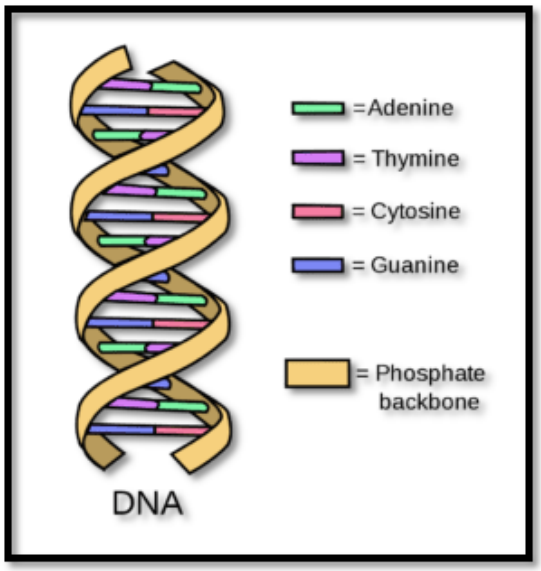

Fig. 1. The DNA structure

DNA was composed of two lengthy strands of nucleotides, forming a double helix structure. A nucleotide comprised a base, a sugar, and a phosphate component. Nucleotides paired up to create base pairs during DNA construction: adenine paired with thymine, and cytosine paired with guanine. Each base was linked to a sugar and a phosphate molecule. In the double helix, base pairs functioned as ladder rungs, while sugar and phosphate molecules formed vertical side rails. DNA's visual representations were available in Figure 1 [1]–[7]. Within cells, 46 elongated structures called chromosomes housed DNA instructions. These chromosomes consisted of smaller DNA segments called genes. Chromosomes had parental origins: one came from the mother and the other from the father.

The creation of a new deep learning model is the paper's major goal and contribution. The Deep DNA Learning Network (DDLN) is its name. When identifying DNAs of people based on their chromosomes, this model is used. After the introduction, the essay is organized as follows: The suggested DDLN model is presented in Section 3, the

Musab T.S. Al-Kaltakchi is with Department of Electrical Engineering, College of Engineering, Mustansiriyah University, Baghdad, Iraq; (e-mail: m.t.s.al_kaltakchi@uomustansiriyah.edu.iq)
Hasan A. Abdulla and Raid Rafi Omar Al-Nima are with Technical Engineering College of Mosul, Northern Technical University, Iraq. (e-mail: raidrafi3@ntu.edu.iq & hasan.alsarraf@ntu.edu.iq).

findings are shown and discussed in Section 4, and the study is concluded in Section 4. Section 2 summarizes earlier work.

II. Prior Work

Systemic Lupus Erythematosus (SLE), an autoimmune condition with variable regional frequency, was the subject of the study in [8]. In North America, there were 23.2 SLE cases per 100,000 people yearly, which was a comparatively larger incidence rate than in Africa, where there were just 0.3 cases per 100,000 people annually. By examining genetic variants that affected Killer T cell activity, the research aimed to foresee the autoimmune response of killer T cells in people with SLE. The study did this by identifying genetic differences using the Boyer-Moore technique and an approximation matching technique. The authors particularly focused on DNA sequences from patients of SLE where the gene nucleotide sequences are connected to T-cell killers in the reference genome of a human. A threshold of a total of 10% length of gene nucleotide sequence was permitted by single nucleotide polymorphisms (SNPs). When 50% of the vulnerability genes was not identical the patient was considered as vulnerable.

in [9] the Unique Personal DNA Pattern (UPDP) was suggested as a new algorithm, where the DNA was used for personal identification. The main aim of this algorithm was to focus on identifying useful repeated DNA patterns. Four databases were employed, namely the DNA Sequences (DS), Human DNA Sequences (HDS), Sample DNA Sequence (SDS), and DNA Classification (DC). The results were such interested, as low errors of False Acceptance Rates (FARs) were benchmarked as 0.75%, 0.26%, 1.41%, and 2.07% for the DS, HDS, SDS, and DC, respectively. Furthermore, there was also the outstanding performance of 0% for each False Rejection Rate (FRR) of any of the four databases.

In [10], it was emphasized that stream-matching methods were employed for the DNA sequencing. In this study, the Rabin-Karp (R-K) and Maximum Common Substream (MCS) algorithms were explored. Different code methods and implementations were assessed. The results provided insightful on the contributed factors that lead to the success in analyzing the DNA sequence by using the R-K and MCS algorithms.

In [11], sensors were used, biometric data were collected and features were extracted. Furthermore, comparisons with templates were considered for the case of identification.

In [12], the development of high-throughput sequencing technology made sequencing individual genomes quick and inexpensive. Compression techniques were required to minimize the burden on data storage and to enable data sharing and administration as the number of sequenced genomes rose. These techniques could be roughly divided into reference-free and reference-based categories. Redundancies in the target DNA sequence were looked for in reference-free approaches to accomplish compression. On the other hand, reference-based techniques found redundant DNA sequences between the target sequence and the reference sequences before compressing them. These techniques worked well with population sequences that were very similar and had few mismatches. Some methods could also be used on sequences with evolutionary links or ones that had some similarities, such as chromosomes. Forensics and DNA applications were presented in [13]. An overview based on human recognition systems was provided in [14].

In the paper [15], the police distributed a picture, but no one claimed the corpse. Eyewitnesses who had earlier seen the man—dubbed the Somerton Man by newspapers—on the beach put his age at approximately 40 and surmised that he was inebriated since they had seen him lift his arm. The authorities treated the death as a suicide even though there was no suicide note because of the location of the death. Blood in the stomach was found during the autopsy, which is frequently indicative of poisoning. Chemical testing, however, was unable to detect any toxicity, probably as a result of the limits of the tools at hand at the time.

In [16], due to processing difficulties and sensitivity concerns, DNA profiling and latent ridge impression analysis (also known as fingerprint analysis) were traditionally treated as independent procedures in forensic investigations. In order to solve this, a novel method was developed that enabled effective DNA profiling and Fingerprint Analysis (FPA) from the same latent material in forensic applications. Features were acquired from both fingerprints and their DNAs. Combination was implemented for inputs to an Improved Artificial Neural Network (IANN). A very high accuracy of 98.54% was successfully achieved for recognition by the network. It suitably predicted whether a user was a client or an impostor. The used method here outperformed other currently exploited approaches. This successfully combine between fingerprint analysis and DNA profiling in terms of forensic research.

In [17], a scalable approach for processing Single Nucleotide Polymorphism (SNP) data of High-Throughput Sequencing (HTS) was described in DNA forensics. The pipeline was used to deal in parallel with numerous samples. This was compatible with both high-performance and standalone computing platforms. Components had been monitored for identifying files of finished sequencing from different platforms like Ion Torrent and Illumina. For automating SNP allele in FASTQ data, GrigoraSNPs software was utilized.

In [18], the understanding of protein-DNA interactions depended on identifying DNA-binding proteins (DBPs). Traditional experimental techniques for identifying DBPs required a significant amount of time and effort. Nevertheless, current methods usually fall short. Recently, machine learning techniques have been examined to sort out such challenges. Restricted Kernel Machines (RKM) of Multi-View

Hypergraph (MVH) was introduced in this work as a new DBP predictor.This method acquired 5 features from 3 protein views. A general hidden vector was utilized for connecting the views. Hypergraph regularization was applied in terms of structural consistency. Results on the PDB186 and PDB1075 databases revealed accuracies of 85.48%and 84.09%, respectively. This surpassed other state-of-the- art techniques. The approach MV-H-RKM was combined with publicly provided code. It also demonstrated promising DBP ability in recognition.

In [19], a new model called the DNA Fine-tuned Language Model (DFLM) was presented. This model exploited human genomics and ChIP-seq Databases. It refined DBP characteristics and acquire sequence dependencies by effectively influencing the human genome information by utilizing different ChIP-seq data. Comparative assessments were evaluated for 69 databases, they showed that the DFLM attained superior performance over other methods. Noticeably, improvement of the DFLM was continued for smaller datasets and more difficult DBPs. Visualization analysis yielded that acquired sequence dependencies of the DFLM were more influential than classical one-hot encoding.

In [20], the discovered that gene demonstration regulation needed DNA motifs recognition. DNA motifs were certain sequence patterns that connected to proteins. The performance of such motifs immediately their lengths were hinged. They are still considered as a challenging aspect for accurately assessing. "MotifLen" was introduced in this work as a new machine learning method, it was designed to obtain motif length and it is a supervised method. This method utilized the convolutional neural network and produced data for assessing aspects. By enhancing the temporal performance of motifs and optimizing them, the proposed approach was modified upon providing motif exploration techniques. Results revealed that MotifLen obtained a percentage accuracy outstanding of 90% for validation. It overcame other compared methods by using real databases. It also effectively improved motif exploration methods. Overall, MotifLen could be considered as a promising issue for accurate motif length assessing for analyzing DNA sequence.

In [21], one of essential steps in DNA ploidy including Image-Based Cytometry (ICM) was the nuclei segmentation. It was crucial for precise estimation of DNA content. Existing supervised methods necessitated laborious pixel-wise labeling. This study introduced an innovative approach: a framework for weakly supervised nuclei segmentation using sparsely annotated bounding boxes instead of segmentation labels. The approach combined fully supervised instance segmentation with self-training and traditional image segmentation. Coarse masks for nucleus-bounding boxes were generated through standard segmentation, followed by mask refinement and pseudo-label generation for unlabeled nuclei using a teacher model. Both teacher and student models shared the same architecture, with the student model initialized from the teacher model. The collaboration of pseudo labels, refined masks, and manual bounding boxes facilitated the training of the student model.

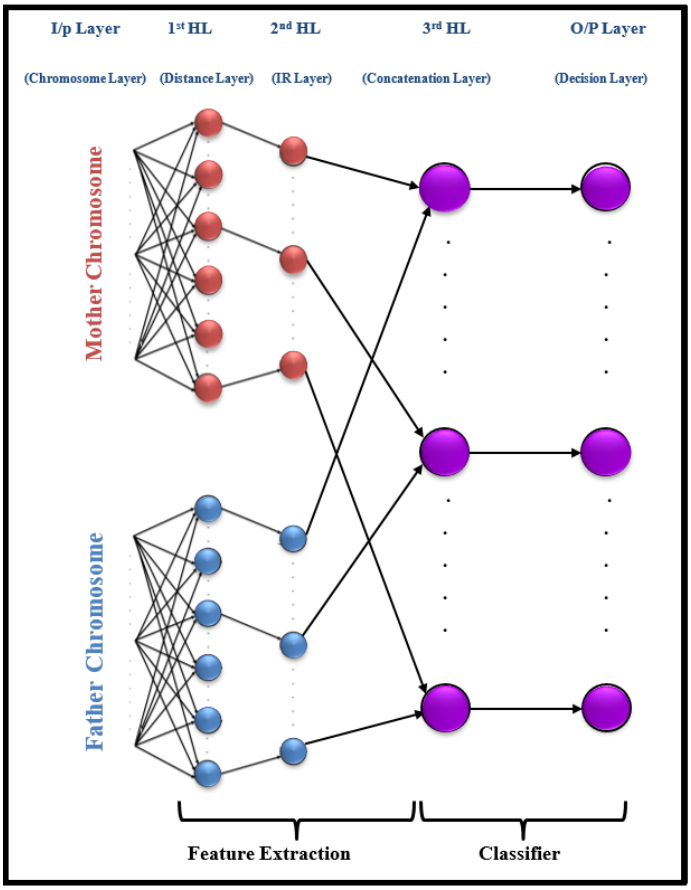

Fig. 2. The framework of the novel DDLN approach

III. Proposed Method

This research introduces a novel deep learning architecture named the Deep DNA Learning Network (DDLN). Comprising five layers, namely the input layer (chromosome layer), the first hidden layer (distance layer), the second hidden layer (Impulse Response (IR) layer), the third layer (concatenation layer), and the output layer (decision layer), this network is designed. The initial two hidden layers serve as a feature extraction component, while the final two layers' function as a classifier.

The structure of this innovative DDLN model is depicted in Figure 2. The input layer accommodates two input vectors, (X_1 andX_2). The first vector, X_1, corresponds to input values from the mother's chromosome, whereas the second vector, X_2, comprises input values from the father's chromosome. The first hidden layer calculates the Euclidean distance between the inputs and assigns weights to each chromosome, as represented by the subsequent equations:

$$y_{i1,j1} = \left\| \mathbf{X_1} - \mathbf{W}_{i1,j1} \right\|, \; i_1 = 1, 2, ..., n_1, \; j_1 = 1, 2, ..., m_1 \tag{1}$$

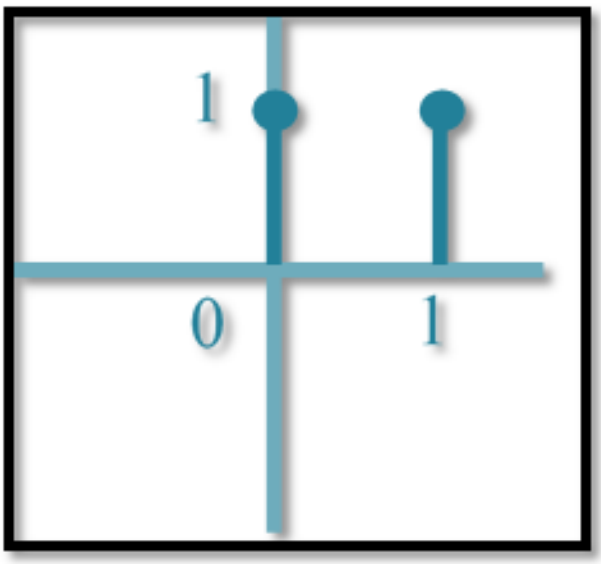

Fig. 3. An illustration of the IR function used in the second hidden layer

$$\mathrm{y}_{i2,j2} = \left\| \mathbf{X_2} - \mathbf{W}_{i\mathbf{2},\mathbf{j2}} \right\|, \; i_2 = 1, 2, ..., n_2, \; j_2 = 1, 2, ..., m_2 \tag{2}$$

In the given context, let's consider the following definitions: $\mathrm{y}_{i1,j1}$ denotes a node value situated within the first hidden layer corresponding to the mother's chromosome. $\mathbf{W}_{i\mathbf{1},\mathbf{j1}}$ represents the weight vector linked to the first hidden layer for the mother's chromosome. Here, n_1 represents the total count of chromosome values pertaining to the mother, while m_1 signifies the quantity of chromosome training vectors associated with the mother. Similarly, $\mathrm{y}_{i2,j2}$ stands for a node value within the first hidden layer related to the father's chromosome, and $\mathbf{W}_{i\mathbf{2},\mathbf{j2}}$ denotes the weight vector affiliated with the first hidden layer for the father's chromosome. In addition, n_2 denotes the total number of chromosome values for the father, and m_2 corresponds to the number of chromosome training vectors for the father.

Within the second hidden layer, an Impulse Response (IR) is computed, labeled as δ_{j1} or δ_{j2}, which is based on the calculated Euclidean distance. Specifically, δ_{j1} signifies the IR associated with the mother's chromosome, while δ_{j2} represents the IR corresponding to the father's chromosome. An illustration exemplifying the application of the IR function within the second hidden layer can be observed in Figure 3.

In reality, the result of the Euclidean distance produces a value that falls within the range of either 0 or 1. These particular values remain within an acceptable threshold, a tolerance that aligns with the parameters set by the employed real Iraqi datasets. The equations governing the computation of the Impulse Response (IR) function can be determined as follows:

$$\delta_{j1} = \begin{Bmatrix} 1 & if \; \mathrm{y}_{i1,j1} = 0 \text{ or } 1 \\ 0 & \text{otherwise} \end{Bmatrix} \tag{3}$$

$$\delta_{j2} = \begin{Bmatrix} 1 & if \; \mathrm{y}_{i2,j2} = 0 \text{ or } 1 \\ 0 & \text{otherwise} \end{Bmatrix} \tag{4}$$

Therefore, the third layer effectively combines the values of δ_{j1} and δ_{j2}. To reiterate, δ_{j1} pertains to the chromosome of the mother, while δ_{j2} corresponds to the chromosome of the father. The equation governing the third layer's operation can be represented as follows:

$$Z_k = \begin{Bmatrix} 11 & if \;\; \delta_{j1} = 1 \; and \; \delta_{j2} = 1 \\ 10 & if \;\; \delta_{j1} = 1 \; and \; \delta_{j2} = 0 \\ 01 & if \;\; \delta_{j1} = 0 \; and \; \delta_{j2} = 1 \end{Bmatrix}, \; k = 1, 2, ..., q \tag{5}$$

In this context, consider the following definitions: Z_k signifies a node value situated within the third hidden layer, which accounts for both chromosomes. The variable q represents the count of chromosome values, applicable to either the mother (n_1) or the father (n_2), with the condition that (n_1) equals (n_2). As a result, the output layer generates the decision results, wherein each decision outcome corresponds to the essential identification value. This value can be calculated utilizing the subsequent equation:

$$R_k = \begin{Bmatrix} 2 & if \;\; Z_k = 11 \\ 1 & if \;\; Z_k = 10 \\ -1 & if \;\; Z_k = 01 \end{Bmatrix}, k = 1, 2, ..., q \tag{6}$$

where: R_k denotes the identification value attributed to the output layer.

It's noteworthy to highlight that the training weights are initially set based on the input training vectors themselves. This very concept of weight initialization is similarly discussed in [22]. This aspect offers notable advantages to the proposed Deep DNA Learning Network (DDLN) model, including the following:

- It exhibits a high level of flexibility, allowing for the addition or removal of hidden and output nodes.
- The training process doesn't necessitate iterative steps, contributing to a rapid training phase.
- It doesn't fall prey to the challenges of local error that other deep learning models, utilizing backpropagation training algorithms, often face.
- It possesses the capability to discern either one or both parents of an individual.

IV. Results and Discussions

A. Datasets Descriptions

This study utilizes two authentic datasets from Iraq: the initial one is labeled as the Real Iraqi Dataset for Kurd (RIDK), and the subsequent one is denoted as the Real Iraqi Dataset for Arab (RIDA). Each individual is characterized by a chromosome containing 30 values, with 15 inherited from the mother and 15 from the father. The RIDK dataset encompasses

chromosome values for 52 individuals, while the RIDA dataset comprises chromosome values for 200 individuals.

Considering the practical considerations of Iraqi forensic medicine, a tolerance of ± 1 for each chromosome value is deemed acceptable. This tolerance is incorporated into the creation of training data. Thus, the augmentation of applying ± 1 to each chromosome value is employed for both datasets. Consequently, the RIDK dataset generates 1,560 training data points, while the RIDA dataset generates 6,000. Conspicuously, , the actual values of both datasets are retained for the testing phase, involving 52 test data for RIDK and 200 for RIDA.

B. DDLN Performances

The evaluation of the DDLN involves assessing its Equal Error Rate (EER) and processing times using the two utilized datasets. Specifically, the EERs and processing times are measured for the complete utilization of the employed RIDK and RIDA datasets. The performance of the DDLN on both datasets in terms of EERs and processing times is presented in Table I. In addition, Figure 4 shows the training and testing samples for both RIDK and RIDA datasets.

TABLE I. DDLN performances of accuracies and times for the employed datasets

Employed Datasets	RIDK dataset	RIDA dataset
Training time	1.24 Sec.	6.19 Sec.
Testing time	0.10 Sec.	1.08 Sec.
EER	0	0

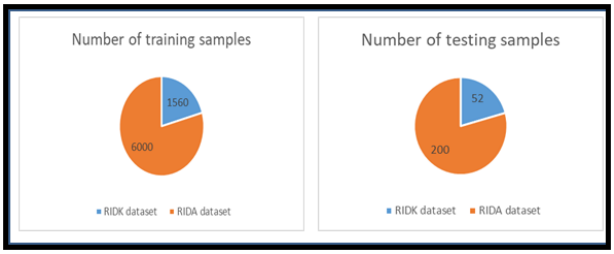

Fig. 4. Number of training and testing samples for RIDK and RIDA datasets

Figures 5 and 6 demonstrate the Receiver Operating Characteristic (ROC) and Detection Error Tradeoff (DET) appearances for our approach. ROC represents the relationship between the False Acceptance Rate (FAR) and True Positive Rate (TPR) which equals to (1 - False Rejection Rate (FRR)). DET represents the relationship between the FAR and FRR [23]–[25]. These figures show the optimal views of ROC and DET. Such appearances are expected for our approach as it obtains the highest accuracy of 100% and lowest Equal Error Rate of 0%.

The outcomes demonstrate the effectiveness of the proposed system with exceptional accuracy rates of 100% achieved for both employed datasets. Furthermore, this study

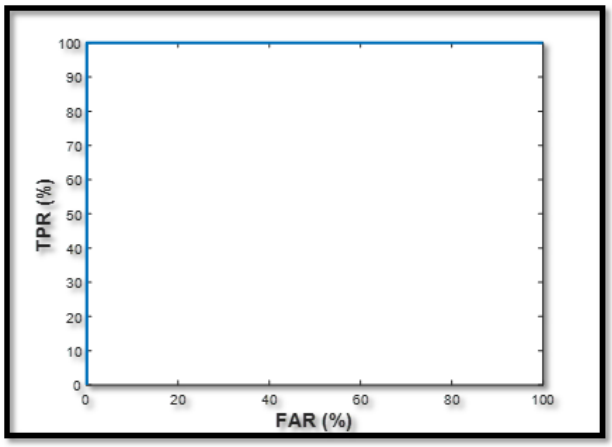

Fig. 5. Demonstration of ROC appearance for our approach

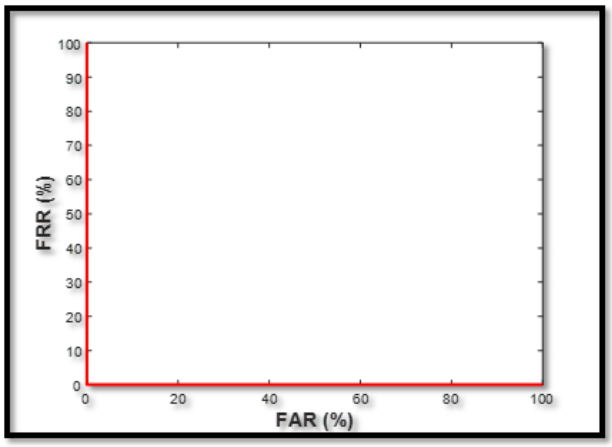

Fig. 6. Demonstration of DET appearance for our approach

considers the time taken for training and testing, and the newly introduced DDLN model exhibits remarkably short durations for these processes. The DDLN's rapid training speed is particularly noteworthy, making it a noteworthy achievement for this approach. In order to validate the reliability of the DDLN, a comparison is made with other deep neural networks.

The DDLN consistently outperforms these models across various aspects, as detailed in Table II. Notably, it excels previous deep learning models of the Stacked Autoencoder [26], Deep Autoencoder Network [27], and Autoencoder Deep Learning [28] in terms of flexibility, training time, Mean Square Error (MSE), and the ability to identify parents. The DDLN's flexibility stands out as it can be adjusted in size without necessitating re-training, unlike other networks that require specific parameters such as hidden layer counts and neurons.

The training times for the DDLN are the shortest among the compared deep learning models for both datasets, presenting a substantial advantage. Although the testing time for

TABLE II. Comparisons between the proposed DDLN and previous deep neural networks (using the same training and testing samples as Table 1)

Deep Learning Model	Parameters	Error and time for the RIDK dataset	Error and time for the RIDA dataset	Identify Parents
SA [26]	NoHL= 3 NoHN in 1st HL = 25 NoHN in 2nd HL = 20 NoHN in 3rd HL = 15	MSE = 0.08 TRT = 6.59 Sec. TET = 0.04 Sec.	MSE = 0.02 TRT = 20.33 Sec. TET = 0.01 Sec.	No
DAN [27]	NoHL= 3 NoHN in 1st HL = 64 NoHN in 2nd HL = 64 NoHN in 3rd HL = 64	MSE = 0.08 TRT = 10.99 Sec. TET = 0.01 Sec.	MSE = 0.02 TRT = 35.67 Sec. TET = 0.02 Sec.	No
ADL [28]	NoHL= 4 NoHN in 1st HL = 30 NoHN in 2nd HL = 30 NoHN in 3rd HL = 30 NoHN in 4th HL = 30	MSE = 0.08 TRT = 8.62 Sec. TET = 0.01 Sec.	MSE = 0.02 TRT = 25.40 Sec. TET = 0.02 Sec.	No
Proposed DDLN	NoHL= 3 NoHN in 1st HL = 2× NoTV NoHN in 2nd HL = 2× NoTV NoHN in 3rd HL = no. of TV vectors	MSE = 0 TRT = 1.24 Sec. TET = 0.10 Sec.	MSE = 0 TRT = 6.19 Sec. TET = 1.08 Sec.	Yes

Where:

- Stacked Autoencoder (SA)
- Deep Autoencoder Network (DAN)
- Autoencoder Deep Learning (ADL)
- NoHL is the number of Hidden Layers.
- NoHN is the number of Hidden Nodes.
- HL is the Hidden Layer.
- TRT is the Training Time.
- TET is the Testing Time.
- NoTV is the number of Training Vectors

the proposed DDLN is slightly longer than some other models, it remains acceptable, particularly for the RIDK dataset. The DDLN's superior performance is further emphasized by achieving the lowest Mean Square Error (MSE) of 0 for both datasets, a feat unmatched by other models. Additionally, the DDLN uniquely possesses the ability to identify the parent or parents of the identified individuals, a feature not present in the other compared deep learning models.

V. Conclusions

A new DDLN model is introduced in this research to identify individuals based on their DNA. This innovative method can ascertain either a single parent or both parents of an individual by utilizing the given chromosomes. The DDLN's adaptability allows for scaling up or down as needed. Unlike other deep learning models, the DDLN's training phase doesn't require iterations and is devoid of local errors. Furthermore, its training process is notably swift. Real datasets from Iraq, referred to as RIDK and RIDA, are employed for evaluation. In both datasets, the proposed method achieves a remarkable EER of 0 and demonstrates superior performance with perfect accuracy of 100%, surpassing other deep learning models.

Acknowledgment

Author thanks to Department of Electrical Engineering, College of Engineering, Mustansiriyah University, Baghdad, Iraq and Technical Engineering College of Mosul, North Technical University, Mosul, Iraq for their constant support, encouragement.

References

[1] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, and D. Zhang, "Biometrics recognition using deep learning: A survey," *Artificial Intelligence Review*, pp. 1–49, 2023. [Online]. Available: https://doi.org/10.1007/s10462-022-10237-x

[2] J. M. Butler, "Recent advances in forensic biology and forensic dna typing: Interpol review 2019–2022," *Forensic Science International: Synergy*, vol. 6, p. 100311, 2023. [Online]. Available: https://doi.org/10.1016/j.fsisyn.2022.100311

[3] T. Sabhanayagam, V. P. Venkatesan, and K. Senthamaraikannan, "A comprehensive survey on various biometric systems," *International Journal of Applied Engineering Research*, vol. 13, no. 5, pp. 2276–2297, 2018. [Online]. Available: https://doi.org/10.1016/j.matpr.2021.07.005

[4] A. K. Singh and A. Mohan, *Handbook of multimedia information security: Techniques and applications*. Springer, 2019. [Online]. Available: https://doi.org/10.1007/978-3-030-15887-3

[5] S. Dargan and M. Kumar, "A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities," *Expert Systems with Applications*, vol. 143, p. 113114, 2020. [Online]. Available: https://doi.org/10.1016/j.eswa.2019.113114

[6] D. H. Fernando, B. Kuhaneswaran, and B. Kumara, "A systematic literature review on the application of blockchain technology in biometric analysis focusing on dna," *Handbook of Research on Technological Advances of Library and Information Science in Industry 5.0*, pp. 77–99, 2023. [Online]. Available: https://doi.org/10.4018/978-1-6684-4755-0.ch005

[7] M. Kumar and A. K. Tiwari, "Computational intelligence techniques for biometric recognition: A review," in *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)*, 2019. [Online]. Available: http://dx.doi.org/10.2139/ssrn.3350260

[8] W. L. Woo, N. O. Ephraim, F. P. Obinna, O. Nwosu, B. O. Jessy, and R. R. Al-Nima, "A scalable algorithm for interpreting dna sequence and predicting the response of killer t-cells in systemic lupus erythematosus patients," *International Journal of Bioinformatics and Intelligent Computing*, vol. 1, no. 1, pp. 48–71, 2022. [Online]. Available: https://www.researchlakejournals.com/index.php/IJBIC/article/view/141/119

[9] M. A. S. Al-Hussein, R. R. O. Al-Nima, and W. L. Woo, "Applying the deoxyribonucleic acid (dna) for people identification," *JOURNAL OF HARBIN INSTITUTE OF TECHNOLOGY*, vol. 54, no. 8, p. 2022, 2022. [Online]. Available: http://doi.org/10.11720/JHIT.54082022.13

[10] S. Tripathi and A. K. Pandey, "Identifying dna sequence by using stream matching techniques," in *2016 International Conference System Modeling & Advancement in Research Trends (SMART)*. IEEE, 2016, pp. 334–338. [Online]. Available: http://doi.org/10.1109/SYSMART.2016.7894545

[11] A. Afolabi and O. Akintaro, "Design of dna based biometric security system for examination conduct," *Journal of Advances in Mathematics and Computer Science*, vol. 23, no. 6, pp. 1–7, 2017. [Online]. Available: http://doi.org/10.9734/JAMCS/2017/27251

[12] B. N.-F. Law, "Application of signal processing for dna sequence compression," *IET Signal Processing*, vol. 13, no. 6, pp. 569–580, 2019. [Online]. Available: https://doi.org/10.1049/iet-spr.2018.5392

[13] D. Primorac and M. Schanfield, *Forensic DNA applications: An interdisciplinary perspective*. CRC Press, 2023. [Online]. Available: https://doi.org/10.4324/9780429019944

[14] D. Palma and P. L. Montessoro, "Biometric-based human recognition systems: an overview," *Recent Advances in Biometrics*, pp. 1–21, 2022. [Online]. Available: https://doi.org/10.5772/intechopen.101686

[15] D. Abbott, "Finding somerton man: How dna, ai facial reconstruction, and sheer grit cracked a 75-year-old cold case," *IEEE Spectrum*, vol. 60, no. 4, pp. 22–30, 2023. [Online]. Available: https://doi.org/10.1109/MSPEC.2023.10092395

[16] J. Johnson, R. Chitra, and A. A. Bamini, "An efficient fingerprint analysis and dna profiling from the same latent evidence for the forensic applications," in *2022 Smart Technologies, Communication and Robotics (STCR)*. IEEE, 2022, pp. 1–6. [Online]. Available: https://doi.org/10.1109/STCR55312.2022.10009376

[17] A. Michaleas, P. Fremont-Smith, C. Lennartz, and D. O. Ricke, "Parallel computing with dna forensics data," in *2022 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE, 2022, pp. 1–7. [Online]. Available: https://doi.org/10.1109/HPEC55821.2022.9926352

[18] S. Guan, Y. Qian, T. Jiang, M. Jiang, Y. Ding, and H. Wu, "Mv-h-rkm: A multiple view-based hypergraph regularized restricted kernel machine for predicting dna-binding proteins," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 2, pp. 1246–1256, 2022. [Online]. Available: https://doi.org/10.1109/TCBB.2022.3183191

[19] Y. He, Q. Zhang, S. Wang, Z. Chen, Z. Cui, Z.-H. Guo, and D.-S. Huang, "Predicting the sequence specificities of dna-binding proteins by dna fine-tuned language model with decaying learning rates," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 1, pp. 616–624, 2022. [Online]. Available: https://doi.org/10.1109/TCBB.2022.3165592

[20] Q. Yu, X. Zhang, Y. Hu, S. Chen, and L. Yang, "A method for predicting dna motif length based on deep learning," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 1, pp. 61–73, 2022. [Online]. Available: https://doi.org/10.1109/TCBB.2022.3158471

[21] Y. Liang, Z. Yin, H. Liu, H. Zeng, J. Wang, J. Liu, and N. Che, "Weakly supervised deep nuclei segmentation with sparsely annotated bounding boxes for dna image cytometry," *IEEE/ACM transactions on computational biology and bioinformatics*, 2021. [Online]. Available: https://doi.org/10.1109/TCBB.2021.3138189

[22] R. R. O. Al-Nima, "Signal processing and machine learning techniques for human verification based on finger textures," Ph.D. dissertation, Newcastle University, 2017. [Online]. Available: https://theses.ncl.ac.uk/jspui/handle/10443/3982

[23] S. Bengio and J. Mariéthoz, "The expected performance curve: a new assessment measure for person authentication," 2003. [Online]. Available: file:///C:/Users/Toshiba/Downloads/rr03-84%20(4).pdf

[24] A. Tharwat, "Classification assessment methods," *Applied computing and informatics*, vol. 17, no. 1, pp. 168–192, 2020. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1016/j.aci.2018.08.003/full/html

[25] Y. Liu and E. Shriberg, "Comparing evaluation metrics for sentence boundary detection," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, vol. 4. IEEE, 2007, pp. IV–185. [Online]. Available: https://doi.org/10.1109/ICASSP.2007.367194

[26] M. Paluszek, S. Thomas, M. Paluszek, and S. Thomas, "Matlab machine learning toolboxes," *Practical MATLAB Deep Learning: A Project-Based Approach*, pp. 25–41, 2020. [Online]. Available: https://doi.org/10.1007/978-1-4842-5124-9_2

[27] A. S. Anaz, M. Y. Al-Ridha, and R. R. O. Al-Nima, "Signal multiple encodings by using autoencoder deep learning," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 1, pp. 435–440, 2023. [Online]. Available: https://doi.org/10.11591/eei.v12i1.4229

[28] A. S. Anaz, R. R. O. Al-Nima, and M. Y. Al-Ridha, "Multi-encryptions system based on autoencoder deep learning network," *Solid State Technology*, vol. 63, no. 6, pp. 3632–3645, 2020. [Online]. Available: http://www.solidstatetechnology.us/index.php/JSST/article/view/3495