RESEARCH PAPER

Autonomous DNA Pattern Intelligence for Large-Scale Forensic Human Identification Using Deep Learning Technique

S. Bavankumar [1*], Dr. V. Rathikarani [2], Dr. R. Santhoshkumar [3]

[1*]Research Scholar, Annamalai University, Chidambaram, Tamil Nadu, India - 608002.
Email: sbavankumar55@gmail.com (Corresponding Author)

[2]Assistant Professor, Annamalai University, Chidambaram, Tamil Nadu, India - 608002.
Email: rathika_1982@rediffmail.com

[3]Associate Professor, Sree Dattha Group of Institutions, Ibrahimpatnam, Telangana, India - 501510.
Email: santhoshkumar.aucse@gmail.com

***Corresponding Author:** S. Bavankumar, Research Scholar, Annamalai University, Chidambaram, Tamil Nadu, India - 608002. Email: sbavankumar55@gmail.com

ABSTRACT

After mass casualty events and in the context of complicated forensic analysis, it is very important to quickly and accurately find missing people for both legal and humanitarian reasons. Short Tandem Repeats (STRs) are the most reliable way to identify a person, but traditional STR analysis is slow and prone to mistakes because it relies on both manual and statistical analysis. This research is especially pertinent in the realm of extensive disasters and tainted DNA specimens. This paper introduces deep learning model convolutional neural network (CNN) based system for the automated analysis of DNA patterns obtained from STR analysis. The STR allele data is changed into a format that deep learning can use. This lets the network learn and find complex DNA patterns in large datasets In accordance with accepted forensic analysis standards, the system is trained and evaluated on both simulated and actual STR analysis. CNN and CNN-Bidirectional models were developed using Label and K-mer encoding techniques to further assess classification accuracy. The tests showed that the K-mer encoding CNN and CNN-Bidirectional models were 98.99% accurate. The new system finds things much faster and with much more accuracy. This makes it useful for things like mass graves and mass casualty events in forensic science. This study improves the accuracy and effectiveness of DNA identification and lays the groundwork for future AI-based forensic analysis.

Keywords: DNA, CNN, STR, Mass Causality Incidents.

How to cite this article: Bavankumar S, Rathikarani V, Santhoshkumar R. Autonomous DNA pattern intelligence for large-scale forensic human identification using deep learning technique. Int J Drug Deliv Technol. 2026;16(3s): 890-898; DOI: 10.25258/ijddt.16.3s.109

1. INTRODUCTION

Forensic DNA profiling is now used a lot more than it was in the mid-1980s when it was first used. It is a well-known way to find criminals from crime scene samples, figure out who the father is, and find the bodies of people who have gone missing. It has become the best way to find victims, especially in mass casualty incidents.

Mass casualty incidents are sudden, unexpected occurrences that negatively affect a community by causing deaths beyond the community's ability to cope with. The etiologies may include natural disasters, warfare, terrorism, and accidents. Notably, the 1990 Scandinavian Star ferry fire was the first mass tragedy to use DNA testing to identify fatalities.

About 1,700 Iraqi army and security personnel were slain by the Islamic State of Iraq and the Levant (ISIL) in Camp Speicher in Tikrit, Iraq, on June 12, 2014. A lot of people died because of this. The examination of human remains post-violent incidents or conflict serves two primary objectives. First, it is very important to find out what caused the death and how it happened during a crime investigation. Second, it is crucial to identify the human remains and, if possible, restore them to their families when the reason and manner of death have been established. The second objective will be the main topic of this essay.

The paper propose deep learning technique for the identification of losing individuals in huge casualty incidents. Rather than the conventional approach of manually juxtaposing the missing individual's profile with those of their relatives. When you do manual matching, you can only compare three to four profiles at a time because it takes a lot of work to look at 15

loci, each with two alleles for one short tandem repeat (STR) profile. Adding an Artificial Intelligence system to find missing people makes the process easier and faster while also making it more effective. The proposed system can also identify samples to find the perons.

In order to determine how close the DNA profiles of missing people are to those of their living relatives their mother, father, and siblings this article suggests a deep learning framework. A DNA short tandem repeat (STR) profile with 15 loci each with two alleles is used in the proposed framework. The missing person's relationship to the reference family should be detectable by the framework. The suggested architecture is effective in developing a deep learning framework to assist in the search for missing persons because it was thoughtfully planned and maintained. In Figure 1, the suggested system is displayed.

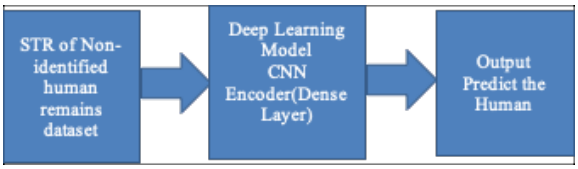

Fig. 1. Missing person identification

2. Litrature Survey

According to Donya A. Khalid and others, DNA profiling is a crucial component of forensic science and aids in the recovery of missing persons following major tragedies. An artificial intelligence system locates victims using DNA Short Tandem Repeat (STR) data. Deep learning models such as the Gated Recurrent Unit (GRU), Bidirectional GRU (Bi-GRU), and Deep Neural Network (DNN) are used to achieve this. When STR information is obtained from living relatives, such as parents or siblings, it can be challenging to identify a victim. Synthetic family trees are created using Al-Najaf Iraqi data. Two members of each gender make up a ten-person household. Because of strict rules and the need to keep family information private, Iraq didn't have many datasets. This project was helpful. It got information from 106 sources about 151,580 people.

Musab T.S. Al-Kaltakchi and others say that DNA is an important physiological biometric because every cell has it. This includes blood, skin, and hair. This study presents the Deep DNA Learning Network (DDLN), an innovative DNA-based personal identification technique. This advanced machine learning model takes DNA from both parents. The model can change shape and size, and it uses the chromosomes it gets to figure out lineage. The DDLN model trains faster than other deep learning methods. The experiment utilizes two authentic Iraqi databases: RIDK for Kurdish speakers and RIDA for Arab speakers. The suggested DDLN model works well on both databases.

Jiany X. Zhang and others Targeted high-throughput DNA sequencing has been utilized to obtain DNA information, alongside genomics and molecular diagnostics. The hybridization kinetics of oligonucleotide probes that cover target gene loci unevenly raise the cost of sequencing. and makes sensitivities lower. This study introduces a deep learning model (DLM) that uses DNA probe sequences to predict NGS depth. To determine the identities of DNA nucleotides and the likelihood that a nucleotide is unpaired, our DLM employs a bidirectional recurrent neural network. Three NGS panels use our DLM: one for human SNPs (39,145 plex), one for human lncRNA (2000 plex), and one for non-human DNA data storage (7373 plex).

3. MISSING PERSON IDENTIFICATION

In the next section, DNA analysis is becoming more and more integral to criminal investigations, especially with regard to missing persons and disaster victims. Short tandem repeats (STRs) are useful markers for DNA analysis used in searching for persons. Even when direct analysis is not possible, the need for assessing familial relationships in their totality is highlighted. The process of DNA-STR profiling is described, from sample procurement to separation and comparison with reference samples. This helps in the identification of persons, resolution of missing persons cases, and determination of paternity. Forensic investigations are increasingly requiring DNA analysis, which is an important tool for finding missing people and identifying victims of disasters. Interpol says that DNA testing, fingerprinting, and dental identification should all be used.

Of the 1153 bodies discovered at Camp Speicher, the Iraqi Ministry of Health reports that over 704 have been identified. But they are still looking at 503 bodies. DNA samples from dead people and living relatives are compared to figure out who they areHowever, there is a significant issue that may be difficult to resolve, and this type of comparison is feasible. The DNA profiles of the victim's living relatives, including their siblings, grandparents, uncles, aunts, cousins, and other relatives, are crucial when the victim's parents are deceased or live far away.

3.1 Forensic identification of humans

Forensic identification entails ascertaining the

specimen's origin by juxtaposing its characteristics with those of a reference sample. DNA markers found in bodily fluids like saliva and blood are often used to identify people. Forensic science usually uses these DNA markers.

Restriction fragment length polymorphism (RFLP) is one method of DNA fingerprinting that looks for differences in the lengths of DNA fragments. This technology had never been used in forensic science before. Genetic analysis shows that people who are related through their mothers have the same mitochondrial DNA (mtDNA). It is possible to look at mtDNA from the mother's side of a victim's brother to help find a missing person or confirm someone's identity.

Y-chromosome analysis is used in forensic science and the study of how humans evolved. Single-nucleotide polymorphism (SNP) is the process of typing for differences in DNA sequences that are caused by changes to a single nucleotide (A, T, C, or G) in the genome.

3.2 DNA–STR

DNA sequences known as Short Tandem Repeats (STRs) are found throughout the genome and range in length from two to six base pairs. These brief periods can be repeated numerous times, though the frequency of repetition varies greatly from person to person. Because of this high degree of heterogeneity, samples may be distinguished from one another. By examining each cell line's distinct STR profile, we can distinguish between them.

Reproducing your research findings may be impossible if you misidentify a cell line, contaminate it, or improperly annotate it. By requiring documentation of the cell lines used in research, the NIH significantly upped the bar for grant applications. Furthermore, cell line identity verification is now required by several journals. A fairly dependable method for locating cell lines that have been misdiagnosed, cross-contaminated, or genetically drifted is WiCell's STR profiling. This implies that you can be certain that the study materials you're using are accurate.

Making sure the cell lines you utilize for research are authentic may be done easily and affordably with WiCell's STR profiling. We can assist you with the creation or verification of a STR profile using the Promega PowerPlex 16HS equipment. This highly sensitive technique can identify one mal/female determination marker (amelogenin) and 15 distinct loci. Additionally, it may detect contamination in other cell lines at as low as 2–5% levels. We create extremely trustworthy data, which is verified by a number of skilled professionals before being reported. Additionally, we will contrast your samples with any other cell lines that WiCell has previously examined from your organization. This will enable us to quickly identify any samples that are incorrectly labeled.

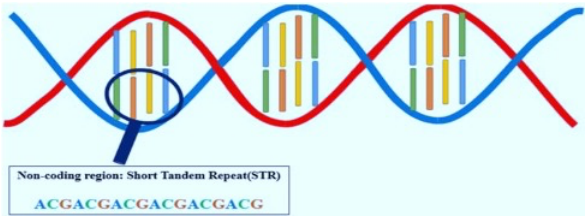

Fig. 2. DNA pair of sequence

4. DEEP LEARNING: DNN & CNN

4.1 DNNs

Deep Neural Networks are a type of artificial neural network that simulates biological neurons by using a structure with multiple layers. This method is based on how the brain works with information. Each of the network's numerous tiers processes and abstracts data differently. Each layer verifies the incoming data before forwarding it to the subsequent layer. The output, or the outcome of the categorization, is provided by the final layer. This helps the network understand more about the data's features and representations, which makes it more accurate and faster than other ways to teach machines. DNNs change the weights of the connections between units so that the predicted outputs are more like the actual outputs. You keep using the backpropagation algorithm until the error is small enough. Deep Neural Networks have been useful for a lot of things, like understanding speech and images, translating languages, and processing natural language (NLP). They are especially good at figuring out how the input and output variables are related in ways that are needed to solve the problem.

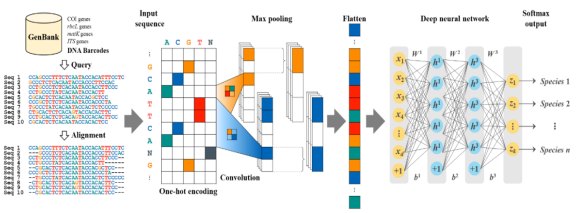

Figure 3: Proposed method of CNN and DNN

It can take a lot of computing power to train deep neural networks (DNNs), and the process can change depending on things like the network's depth and the learning rate. Overfitting can happen if the training data is too small or not very varied. AI is very important in forensic science because it can help different forensic

methods work better and more accurately. AI can quickly look at big sets of data, do boring tasks for you, and find patterns that people might not see. Keep in mind that AI can't replace the knowledge and skills of people. Deep learning is a kind of AI and machine learning that uses multilayer neural networks to find patterns in data that are hard to see. Deep learning's neural networks look at big sets of data on their own to find patterns and links. This technique has been effectively employed in various bioinformatics applications to enhance classification accuracy. Many studies show that deep learning is better than older methods like random forest or support vector machine algorithms at predicting protein binding and how easy it is for DNA sequences to get to. There is a lack of research on using deep learning to find missing people by matching their DNA. However, two relevant studies utilize artificial intelligence techniques in DNA analysis: Anggreainy et al. Employ fuzzy inference for STR-DNA matching across 16 loci, while Siino and Sears utilize gradient descent logistic regression for kinship analysis based on 13 loci.

This paper examines and contrasts three deep learning methodologies for an artificial intelligence system utilizing DNA-STR data from five loci to identify victims: Deep Neural Networks and Convolutional Neural Networks.

$$Z = (\sum X_2 * W_t + b) \qquad \text{--- (1)}$$

In this context, X_t denotes the input vector, W_t represents the weight vector, b indicates the bias, Z signifies the output from the network, and σ refers to the activation function, which may be a leaky rectified linear unit (Leaky ReLU) in all hidden layers or a sigmoid function in the output layer.

Leaky–ReLU Function:

$$(x) = \max(ax, x)$$

----(2)

Sigmoid:

$$a_s(x) = \frac{1}{1+e^{-x}}$$

---(3)

CNN (Convolutional Neural Network) algorithm

1. Convolution Operation

Convolution extracts features (edges, textures, patterns).

For an input image I and a filter/kernel K of size $f \times f$:

$$O(i,j) = \sum_{m=0}^{f-1} \sum_{n=0}^{f-1} I(i+m, j+n) \cdot K(m,n)$$

Where as:

- The output feature map value at (i,j) is O(i,j).
- III = input image
- KKK = convolution kernel

Output Size Formula

If

- WWW = input width
- KKK = kernel size
- SSS = stride
- PPP = padding

Then output width:

$$W_{out} = \frac{W - K + 2P}{S} + 1$$

2. Activation Function (ReLU)

ReLU introduces non-linearity.

Formula

----(6) $x)$

Other activations:

- Sigmoid: $\sigma(x) = \frac{1}{1+e^{-x}}$
- Tanh: $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

3. Pooling Operation

Pooling reduces spatial dimension (downsampling).

Max Pooling Formula

$$P(i,j) = \max_{(m,n) \in R_{ij}} O(m,n)$$

ıe region defined by pooling window.

4. Flattening

Turn the pooling feature map into a one-dimensional vector.

5. Fully-Connected Layer

Each neurons connected with each previous layers.

$$z = W^T x + b$$

Where:

- W = weight matrix
- x = input vector
- b = bias

7. The Loss Function for Cross-Entropy

Taking into account the real label y and the expected probability y:

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

Step	Formula
Convolution	$O(i,j) = \sum I \cdot K$
Output size	$\frac{W-K+2P}{S} + 1$
ReLU	$\max(0, x)$
Pooling	$\max()/\text{avg}()$ over region
Fully Connected	$z = W^T x + b$
Softmax	$\frac{e^{y_i}}{\sum e^{y_j}}$
Loss	$-\sum y \log \hat{y}$
Weight Update	$W - \eta \frac{\partial L}{\partial W}$

5. Experiments and Results

5.1. Datasets

There are many different types of human DNA identification datasets. Some are large, publicly available reference genomes (like GRCh38), while others are specialized forensic databases used to match DNA profiles. The NIH's GenBank and dbSNP for genetic variation, the CODIS system for forensic profiles, and research-oriented datasets on sites like Kaggle for machine learning applications are all important resources.

For example: NCBI Databases (GenBank, dbSNP, dbVar): The main place to find DNA sequences that are open to the public, including SNP data for analyzing variation. The DNA Classification Dataset (Kaggle) is a fake dataset with 3,000 samples and 13 features (like GC content and sequence length) that can be used to sort DNA sequences.

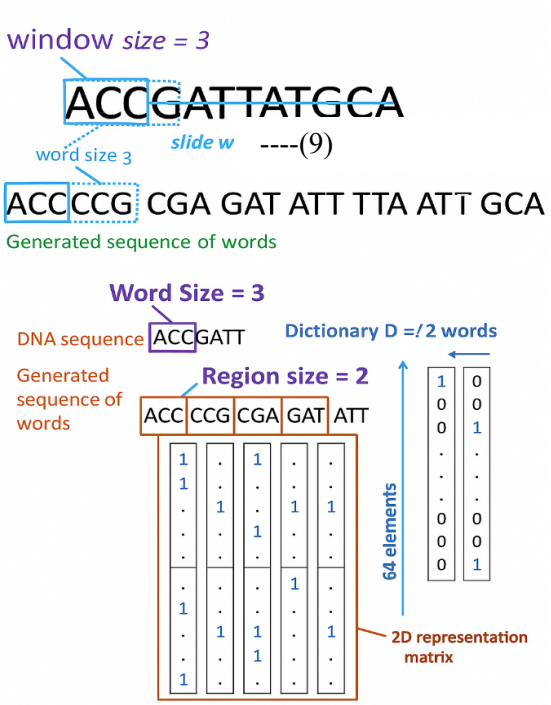

Figure 4. Changing a DNA sequence into a string of words and a one-hot vector that shows the DNA sequence.

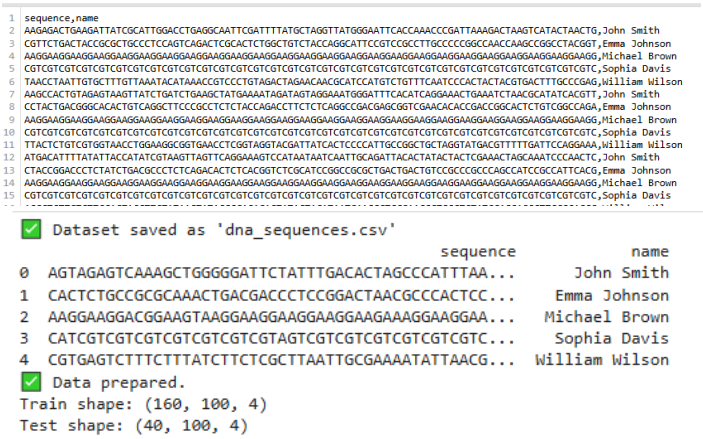

Figure 5. Screenshot of Dataset Description

- **Training Set:** 160 samples, each with a shape of (100, 4), indicating 100 positions in the sequence and 4 possible nucleotide channels (A, T, C, G) per position—likely one-hot encoded.
- **Test Set:** 40 samples, also with shape (100, 4).

This output shows how workflows in bioinformatics research use DNA sequence data as input for deep learning models.

5.2. Outputs and Discussion

```
model = Sequential([
    Conv1D(64, 8, activation='relu', input_shape=(100, 4)),
    MaxPooling1D(4),
    Dropout(0.3),
    Conv1D(128, 5, activation='relu'),
    MaxPooling1D(4),
    Dropout(0.3),
    Flatten(),
    Dense(128, activation='relu'),
    Dropout(0.4),
    Dense(len(names), activation='softmax')
])
```

Figure 6. Applying Convolution Neural Network

You can make a linear stack of neural network layers with the Sequential() model. The output of each layer is the input for the next layer. The first layer is a Conv1D layer with 8 kernels and 64 filters. It looks for short patterns and motifs in the DNA sequence that are important for biology. The input shape is 100 bases long and has four dimensions. Each base (A, C, G, and T) is represented by a one-hot vector. The next layer is MaxPooling1D, and it has a pool size of 4.

This makes the sequence simpler while keeping the most important parts, which makes the computer run better. Dropout (0.3) turns off a random 30% of the neurons during training to keep the model from overfitting. The second Conv1D layer has 128 filters, but the kernel size is only 5. This lets the network learn more complex and higher-level sequence features, which makes the model deeper. We then use MaxPooling1D and Dropout (0.3) to keep things normal and cut down on the number of features. The Flatten layer turns the extracted feature maps into a one-dimensional vector so that fully connected layers can use them. We can learn more about how features are connected by turning on a Dense layer with 128 neurons and ReLU. To keep the model from overfitting, a higher Dropout rate of 0.4 is used before the output layer.

One neuron for each person (class) and a softmax activation make up the final Dense layer. Probability scores for person identification are provided by this. After 16 batches of 20 epochs, the model is trained using X_train (DNA sequences) and y_train (one-hot encoded identities). A 10% validation split is used to keep an eye on how well the model works in general. Verbose=1 shows that the training is going well. This code builds a 1D convolutional neural network (CNN) to sort DNA sequences into people. It also handles training and validation and has a strong prediction function that checks the input's validity and length. This kind of automation is what makes forensic and biological pipelines work well and be easy to use. Finally, given a test sequence like

```
test_sequences = [
    "AGCTAAACAGCTAAACAGCTAAACAGCTAAACAGCTAAACAGCTAAACAGCTAAACAGCTAAACAGCTAAAC",
    "CCGGTTCCGGTTCCGGTTCCGGTTCCGGTTCCGGTTCCGGTTCCGGTTCCGGTTCCGGTTCCGGTTCCGGTTCC",
    "AAAAGGGGAAAAGGGGAAAAGGGGAAAAGGGGAAAAGGGGAAAAGGGGAAAAGGGGAAAAGGGGAAAAGGGG",
    "CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG"
]
```

Figure 7. Screenshot of Testing data for prediction

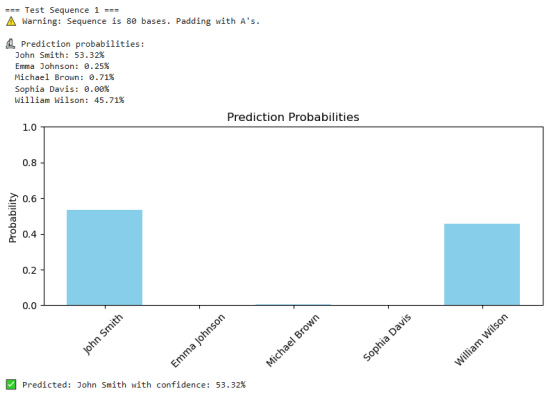

Figure 8. The first test data it predicted John smith more confidence

With a maximum probability of 53.32%, the model indicates that "John Smith" is the most probable first sequence match.

The model got a DNA sequence that wasn't very long, figured out how long it was on its own, and made a good guess that "John Smith" was the most likely person. It also showed how likely each of the five contenders was to win.

.

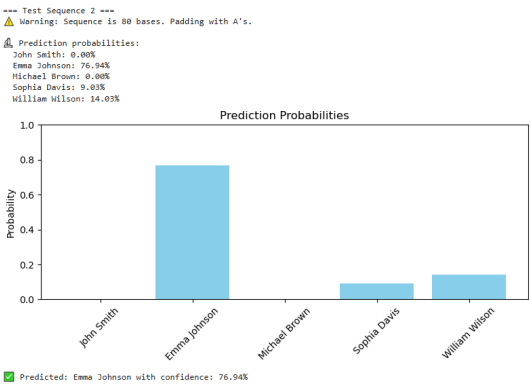

Figure 9. The second test data it predicted Emma Jhonson more confidence

The model is 76.94% sure that "Emma Johnson" is the best match.

The numbers and bar chart show that the model confidently matched "Emma Johnson" to a padded DNA sequence.

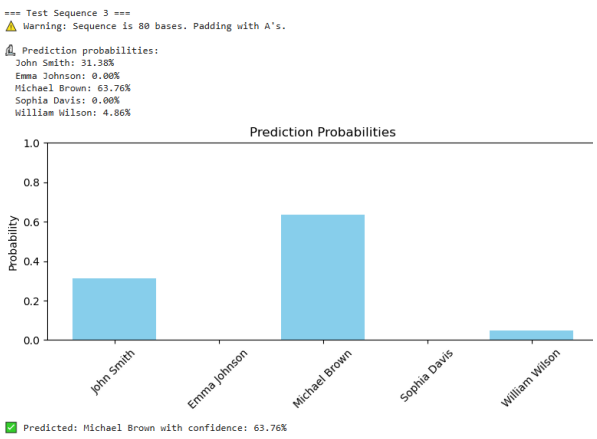

Figure 10. The Third test data it predicted Michael Brown more confidence

The model is 63.76% sure that "Michael Brown" is the best match.

Both the numerical results and the bar graph demonstrate that the model had high confidence in its identification of "Michael Brown" as a match when given a padded DNA sequence.

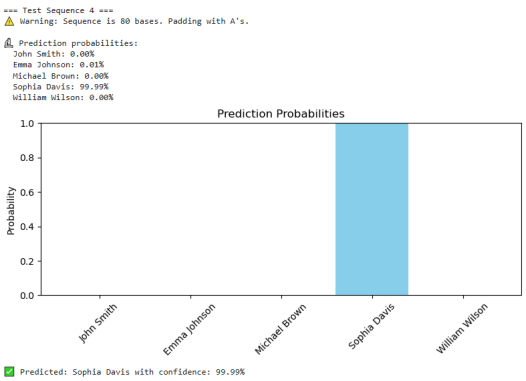

Figure 11. The fourth test data it predicted Sophia davis more confidence

The model says that "Sophia Davis" is the most likely match, with a 99.99% chance. The model got a DNA sequence that was not very long, changed its length on its own, and made a good guess that "Sophia Davis" was the most likely candidate. It showed the chances of each of the five choices next to the guess.

5.3 Metrics for evaluating performance

This study employs the following metrics to compare and analyze the suggested deep learning model's efficiency:

The ratio of correct predictions to total predictions is what defines prediction accuracy.

$$Accuarcy = \frac{TP + TN}{TP + TN + FP + FN}$$

Sensitivity (Recall): It is the true positive-to-actual positive ratio.

$$Sensitivity = \frac{TP}{TP + FN}$$

Precision: The accuracy of the model is measured by the ratio of correct predictions to all positive predictions.

$$Precison = \frac{TP}{TP + FP}$$

F1_Score: It strikes a balance between being sensitive and being precise. It was once a standard for gauging how well the test performed on the positive class.

$$F1 - Score = 2 * \frac{Sensitivity * Precision}{Sensitivity + Precision}$$

A receiver operating characteristic (ROC) curve is a graph that shows how well a classification model works at different thresholds. The ROC curve shows the true positive rate and the false positive rate. The Area under Curve (AUC) is a single ROC curve statistic that measures how well a binary classifier can separate classes. To check the predictions, four metrics are used: TP, FP, TN, and FN.

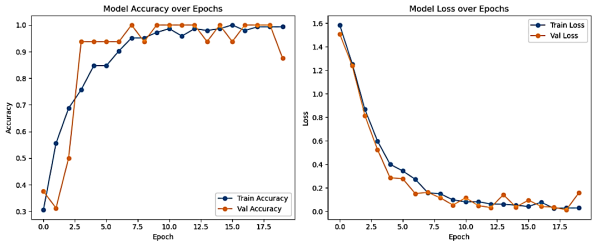

Figure 12. Model accuracy and loss over epochs

The blue line ("Train Accuracy") shows how well the model did on the training data. The training accuracy reaches a high point of 98.99% starting with epoch 7. This means that the model is able to correctly predict training instances. The orange line ("Val Accuracy") shows how well the model did on the new validation data. This is a more accurate way to measure how well the model works in general. The validation accuracy reaches a high of 98.99% at different times (epochs 10–12 and 15–17). The last epoch, Epoch 19, sees a drop to 0.88 (88%). As shown in the picture, the model got and kept an amazing accuracy of 98.99% for both the training and validation sets.

6. CONCLUSION:

CNN-based analysis of the pattern of DNA proves to be of tremendous value. This is especially so when it comes to cases of missing persons, mass casualties, and other forensic cases. For one, CNN-based analysis

ensures swift and precise handling of large volumes of information relating to STR patterns. This, in turn, means faster and more accurate identification of individuals, thus ensuring swift recovery of victims in disaster cases, as well as modernized means of solving crimes. By mixing two distinct approaches in data analysis, criminals and law officials are provided with stronger and more objective means of analyzing forensic data. For instance, the validation accuracy peaks at 98.99%, and in training data, the accuracy of results obtained was optimal. At the end of training, there is, however, a slight drop in validation accuracy, which can be attributed to the generalization capabilities of the network.

References:

[1] Eickholt, J. and Cheng, J. (2013) DNdisorder: Predicting Protein Disorder Using Boosting and Deep Networks. BMC Bioinformatics, 14, 88-98. http://dx.doi.org/10.1186/1471-2105-14-88

[2] Leung, M.K.K., Xiong, H.Y., Lee, L.J. and Frey, B.J. (2014) Deep Learning of the Tissue-Regulated Splicing Code. Bioinformatics, 30, i121-i129. http://dx.doi.org/10.1093/bioinformatics/btu277

[3] Lee, H., Grosse, R., Ranganath, R. and Ng, Y.A. (2009) Convolutional Deep Belief Networks for Scalable Unsuper vised Learning of Hierarchical. Proceeding of the 26th Annual International Conference on Machine Learning, Mon treal, 14-18 June 2009, 609-616. http://dx.doi.org/10.1145/1553374.1553453

[4] Mikolov, T., Chen, K., Greg, C. and Dean, J. (2013) Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.

[5] Johnson, R. and Zhang, T. (2015) Effective Use of Word Order for Text Categorization with Convolutional Neural Networks. Proceeding of Human Language Technologies: The 2015 Annual Conference of the North American, Den ver Colorado, 31 May-5 June 2015, 103-112. http://dx.doi.org/10.3115/v1/n15-1011

[6] Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D.K. and Young, R.A. (2005) Genome-Wide Map of Nucleo some Acetylation and Methylation in Yeast. Cell, 122, 517-527. http://dx.doi.org/10.1016/j.cell.2005.06.026

[7] Higashihara, M., Rebolledo-Mendez, J.D., Yamada, Y. and Satou, K. (2008) Application of a Feature Selection Me thod to Nucleosome Data: Accuracy Improvement and Comparison with Other Methods. WSEAS Transactions on Bi ology and Biomedicine, 5, 153-162.

[8] Li, J. and Wong, L. (2003) Using Rules to Analyse Bio-Medical Data: A Comparison between C4.5 and PCL. Pro ceedings of Advances in Web-Age Information Management 4th International Conference, Chengdu, 17-19 August 2003, 254-265. http://dx.doi.org/10.1007/978-3-540-45160-0_25

[9] Towell, G., Shavlik, J. and Noordewier, M. (1990) Refinement of Approximate Domain Theories by Knowledge- Based Artificial Neural Networks. Proceedings of the 8th National Conference on Artificial Intelligence, Boston, 29 July-3 August 1990, 861-866.

[10] Butler, J.M. (2015). The future of forensic DNA analysis. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1674): 20140252. https://doi.org/10.1098/rstb.2014.0252

[11] **S. Bavankumar**, Dr. V. Rathikarani, Dr. R. Santhoshkumar, "REVOLUTIONIZING DNA SPECIES CLASSIFICATIONUSING CONVOLUTIONAL NEURAL NETWORK"International Journal of Applied Mathematics, ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version), Vol. 38 No. 10s 2025, Pg. No: 795-806.

[12] **S. Bavankumar**, Dr. V. Rathikarani, Dr. R. Santhoshkumar, "Cnn-based Dna Pattern Analysis for Missing Person Identification in Mass Casualty and Forensic Scenarios", Pre Print on Research Square, Under Peer Review on *International Journal of Computational Intelligence Systems*.

[13] **S. Bavankumar** et al." Detecting and Responding to Cloud Security Incidents based on AI and Forensic Approach" 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Date of Conference: 14-15 December 2023, IEEE Xplore.

[14] B Rajalingam, R Santhosh Kumar, P Deepan, P Santosh Kumar Patra, **S Bavankumar,** "A Smart System for Sign Language Recognition using Machine Learning Models", 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), IEEE, Page no: 1125-1131, (2022).

[15] **S. Bavankumar,** Dr. B. Rajalingam, Dr. R. Santhoshkumar, Dr. G. JawaherlalNehru, P.Deepan, "A Real Time Prediction and Classification of Face Mask Detection using CNN Model", Turkish Online Journal of Qualitative Inquiry, Vol. 12, Issue: 8, (2021).

[16] Mintu Debnath, Vijay Kumar Salvia, Alighazi Siddiqui, Khalid Ali Qidwai, P. Durga Devi, **S. Bavankumar**, "Mathematical Foundations of

Explainable AI: A Framework based on Topological Data Analysis", Communications on Applied Nonlinear Analysis, Vol. 32 No. 9s (2025)**.**

[17] Dr. R. Santhoshkumar, Dr. B. Rajalingam, Dr. P. Deepan, Dr.G. JawaherlalNehru, **S. Bavankumar**" Plant Leaf Disease Prediction: A PLDD Net-SVM Model Proposed using Internet of Thing (IOT) and Integrated Learning Model" Journal of optoelctronics Laser, Vol. 41 No. 9 (2022).

[18] **S.Bavankumar,** Dr.B.Rajalingam, Dr.R.Santhoshkumar" A real Time Prediction and Classification of Face Mask Detection using CNN Model" Turkish Online Journal of Qualitative Inquiry(TOJQI), Vol. 12 No.8 (2021).

[19] Akash Tomar, Anil Pimplapure, **S Bavankumar,** M Prathyusha, Anil Kumar Sahu, Sourabh Kumar Jain, "An Empirical Investigation of the Impact of Data-Driven Decision-Making on Supply Chain Resilience in the Automotive Industry", Tuijin Jishu/Journal of Propulsion Technology, Vol. 44, Issue:3, Page no:1-7(2023).

[20] Dr. R.Santhoshkumar Dr. G.JawaherlalNehru, Dr. S.Jothilakshmi, Dr. S. Jothishri, **S. Bavankumar,**Dr. B. Rajalingam, "Music Genre Classification Using Deep Learning Techniques", Turkish Online Journal of Qualitative Inquiry, Vol. 12 No.8 (2021).

[21] Dr. P. Santosh Kumar Patra Dr. B.Rajalingam, Dr. R.Santhoshkumar, Dr. G.Govinda Rajulu, P.Deepan, **S. Bavankumar**,"Hybrid Integration of Transforms with Neural Network based Fusion Techniques for clinical and Healthcare Applications" Turkish Online Journal of Qualitative Inquiry(TOJQI), Vol. 12 No.9 (2021).

[22] P. Mahesh,Dr. K. Srinivas,Dr. Santhoshkumar,Dr. B. Rajalingam,Dr. JawaherlalNehru, Dr. M. Vadivukarassi, **S. Bavankumar,** "Review of Security Threats in Cloud Computing and its Protection" Res Militaris, Volume -14 , Issue -4 : Page: 1-7 (2024).

[23] K Nagaraju,Dr. K. Srinivas,Dr. Santhoshkumar,Dr. B. Rajalingam , Dr. JawaherlalNehru, Dr. M. Vadivukarassi, **S. Bavankumar**, "FUNCTION OF CLOUD COMPUTING IN SMART AGRICULTURE", Res Militaris, Volume -14 , Issue -4 : Page: 8-10

[24] Gunda Satish Kumar,Dr. K. Srinivas,Dr. Santhoshkumar,Dr. B. Rajalingam , Dr. JawaherlalNehru, Dr. M. Vadivukarassi, **S. Bavankumar,** "HUMAN COMPUTER INTERACTION IN DAILY LIFE USING DEEP LEARNINIG", Res Militaris, Volume -14 , Issue -4 : Page: 11-18,(2024).

[25] P. Naga Triveni, Dr. G. JawaherlalNehru,Dr. R. Santhoshkumar, **S. BavanKumar,** "Verification and Validation of Certificate UsingBlockchain", Res Militaris, Volume -13 , Issue -4 : Page: 974-978,(2023).

[26] P. Naga Triveni,Dr. G. JawaherlalNehru,Dr. R. Santhoshkumar, **S. BavanKumar,** "A REVIEW ON RAINFALL PREDICTIONUSING MACHINE LEARNING ALGORITHMS: MLR AND ARTIFICIAL NEURAL NETWORK", Res Militaris, Volume -13 , Issue -4 : Page: 979-983,(2023).

[27] K. Priti,Dr. G. JawaherlalNehru,Dr. R. Santhoshkumar, **S. BavanKumar,** "SMART TRAFFIC DETECTION SYSTEM USING CANNY EDGE DETECTION", Res Militaris, Volume -13 , Issue -4 : Page: 984-995,(2023).

[28] G. Kishore, Dr. G. JawaherlalNehru,Dr. R. Santhoshkumar, **S. BavanKumar,** "A NOVEL APPROACH OF FILTERING HASHINSTAGRAM TAGS USING HITS ALGORITHM", Res Militaris, Volume -13 , Issue -4 : Page: 996-1009,(2023).

[29] M Dileep kumar, Dr. G. JawaherlalNehru, Dr. R. Santhoshkumar, **S. BavanKumar,** "Predict Plant Growth and Yield in Greenhouse Environments Using Deep Learning", Res Militaris, Volume -13 , Issue -4 : Page: 1010-1017,(2023).

[30] Venkatarangan J, Dr. G. JawaherlalNehru, Dr. R. Santhoshkumar, **S. BavanKumar,** "Bird Species Identification UsingDeep Learning Techniques", Res Militaris, Volume -13 , Issue -4 : Page: 1018-1023,(2023).

[31] P. Chaitanya,Dr. G. JawaherlalNehru,Dr. R. Santhoshkumar, **S. BavanKumar,** "An intelligent paper currency recognition system", Res Militaris, Volume -13 , Issue -4 : Page: 1024-1030,(2023).