World Academy of Science, Engineering and Technology
International Journal of Biotechnology and Bioengineering
Vol:18, No:10, 2024

Personalized Applications for Advanced Healthcare through AI-ML and Blockchain

Anuja Vyas, Aikel Indurkhya, Hari Krishna Garg

***Abstract*—**Nearly 25 years have passed since the landmark publication of the Human Genome Project, yet scientists have only begun to scratch the surface of its potential benefits. To bridge this gap, a personalized genomic application has been envisioned as a transformative tool accessible to people worldwide. This innovative solution proposes an integrated framework combining blockchain technology, genome-specific applications, and data compression techniques, ensuring operations to be swift, secure, transparent, and space-efficient. The software harnesses advanced Artificial Intelligence and Machine Learning methodologies, such as neural networks, evaluation matrices, fuzzy logic, and expert systems, to analyze individual genomic data. It generates personalized reports by comparing a user's genome with a reference genome, highlighting significant differences. Blockchain technology, with its inherent security, encryption, and immutability features, is leveraged for robust data transport and storage. In addition, a 'Data Abbreviation' technique ensures that genetic data and reports occupy minimal space. This integrated approach promises to be a significant leap forward, potentially transforming human health and well-being on a global scale.

***Keywords*—**Artificial intelligence in genomics, blockchain technology, data abbreviation, data compression, data security in genomics, data storage, expert systems, fuzzy logic, genome applications, genomic data analysis, human genome project, neural networks, personalized genomics.

I. Introduction

THE Human Genome Project (HGP), an unprecedented collaboration of biology and technology, has revealed that human genome contains about 35,000 genes, but only around 20,000 of these are responsible for making proteins [1]. According to the latest updates from the GenCode database (release 45), there are approximately 19,395 protein-encoding genes [2]. The human genome, comprising about 3 billion DNA constructs, was thoroughly mapped by the HGP, resulting in the creation of a 'reference genome' derived from diverse human and model organism sequences.

Despite over two decades of work, the goal of using genome sequencing to provide individuals with meaningful health insights remains unfulfilled. To fill this gap, the present study has been undertaken in which an attempt has been made to develop a personalized genomic application, 'Your Genome App'. This app leverages advanced artificial intelligence (AI) and machine learning (ML) algorithms to precisely analyze individual genomes, detect anomalies, and generate detailed reports with actionable recommendations. Users can gain insights into their physical traits, health status, and disease risk, facilitating necessary follow-ups and early disease prognosis. Further, the app streamlines the sharing of this vital information with healthcare professionals, enabling prompt and effective interventions.

To ensure the security and integrity of patient data, blockchain technology is employed. This technology offers distributed, immutable, and transparent data storage. Smart contracts within the blockchain ecosystem further enhance efficiency by streamlining processes and ensuring faster transactions.

Given the vast amounts of data generated by human genome sequencing, alternative data storage solutions, such as DNA storage [3], [4] have been also explored. However, this method has its challenges, particularly in data retrieval, thus, a data abbreviation approach to address these issues has been devised within this research. By using algorithms to condense repetitive sequences, large genomic files can be more efficiently stored, improving the scalability and accessibility.

II. Materials and Methods

A. Reference Genome

Since the completion of the HGP, numerous enhanced versions of the reference genome have emerged, with the latest being "GRCh38 [5]." The ‘Your Genome App’ operates on the premise that this reference genome is highly accurate and free of errors.

B. Technologies

The present research employs a diverse range of cutting-edge technologies to enhance the analysis, storage, and sharing of genomic data:

- *Machine Learning:* a subfield of artificial intelligence, develops algorithms that can learn from and adapt to new data without explicit programming. Deep learning, a subset of machine learning, allows machines to understand complex patterns and perform intricate tasks. In this research, machine learning is used to process and compare the user's genome with the reference genome.
- Evaluation metrices: such as F1 score, ROC-AUC, accuracy, and precision are utilized to gain insights into different aspects of model performance.

Anuja Vyas is with Institute for Excellence in Higher Education, Bhopal – 462016 India (phone: 91-797-475-1896; e-mail: anujavyas20@gmail.com).

Aikel Indurkhya is with Institute for Excellence in Higher Education, Bhopal – 462016 India (phone: 91-930-275-0880; e-mail: aikelindu02@gmail.com).

Hari Krishna Garg is with Institute for Excellence in Higher Education, Government of Madhya Pradesh, Bhopal – 462016 India (e-mail: drhkgarg@gmail.com).

- Fuzzy Logic and Expert System: advanced AI tools. Fuzzy logic mimics human thought processes to support decision-making by handling data that include various possibilities, such as partial truths and uncertainties. This type of reasoning aids in managing the complexities and variability inherent in genomic data whereas an expert system, using artificial intelligence, replicates the decision-making abilities of a specialist in a specific domain. In this research, the expert system's extensive knowledge of genes and predicted phenotypes helps generate accurate outputs, assisting in the analysis and interpretation of genomic data.
- Blockchain: particularly Ethereum, utilized to create a secure and efficient platform for storing, analyzing, and sharing genomic data. Blockchain ensures data security, immutability, and tamper-proof storage through a decentralized network.
- Encryption: Cryptography, employing the AES 256-bit key algorithm and the RSA Algorithm, ensuring data security through encryption. The private key is used for signing and decryption, while the public key handles verification and encryption. AES-256 encrypts the data, and a separate asymmetric algorithm securely shares the secret key for AES-256.

III. Personalized Genomic Application: 'Your Genome'

The HGP has sparked expectation that soon, individuals will easily access their genome sequences via mobile devices, akin to medical reports. These genomes harbor potential for tailored services, with Personalized Genomic Applications (PGAs) at the forefront. These applications aim for user-friendly interfaces, integrating AI and advanced data tools to compare and assess DNA against reference genomes.

By emphasizing cross-platform development, users can maximize their genetic data benefits. The envisioned application will unravel insights from genomic analysis, predicting disease risks and severity. Anchored on a robust reference genome, algorithms will scrutinize genetic variations, offering personalized insights and proactive health recommendations.

A. Sequencing

Advancements in healthcare technology now empower patients to have their genomes sequenced at specialized facilities, with the capability to securely share these genetic profiles using front-line blockchain technology. Before sequencing takes place, obtaining informed consent from patients is a fundamental prerequisite. Strict adherence to ethical and legal standards is mandatory for laboratories, data collection points, and storage facilities involved in this process.

During sequencing, laboratories produce raw data files (FASTA files) containing millions or even billions of short DNA sequences. Following sequencing, patients have the option to directly store this data on their IPFS (InterPlanetary File System) or on a blockchain. Utilizing IPFS ensures data dispersal across a decentralized network of computers, enhancing resilience and censorship resistance compared to traditional cloud storage methods. Balancing security, privacy, and accessibility becomes paramount in this context.

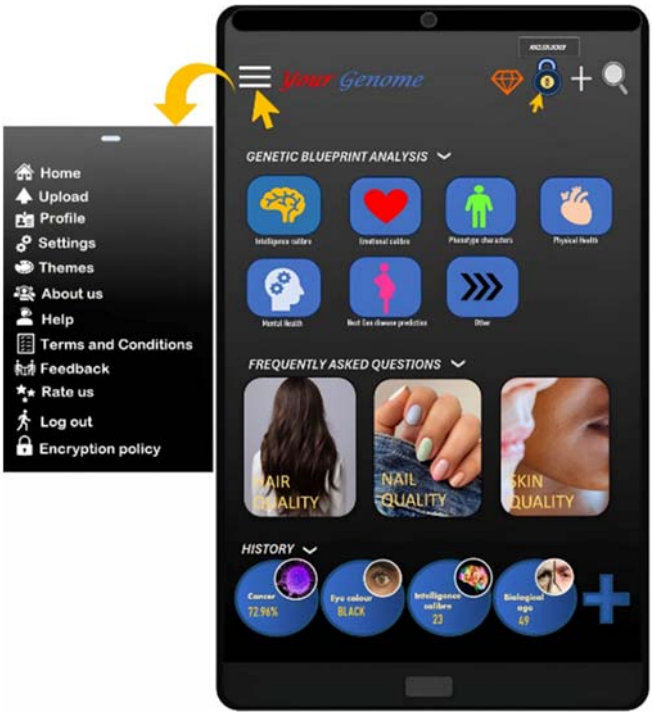

Fig. 1 Notional representation of the home page of the application – 'Your Genome'

Upon processing and preparing genomic data for storage, IPFS generates a unique Content Identifier (CID), akin to a blockchain hash, which serves as an exclusive marker for the data stored within IPFS. This approach marks a significant step forward in securely managing and leveraging genetic information for healthcare advancements.

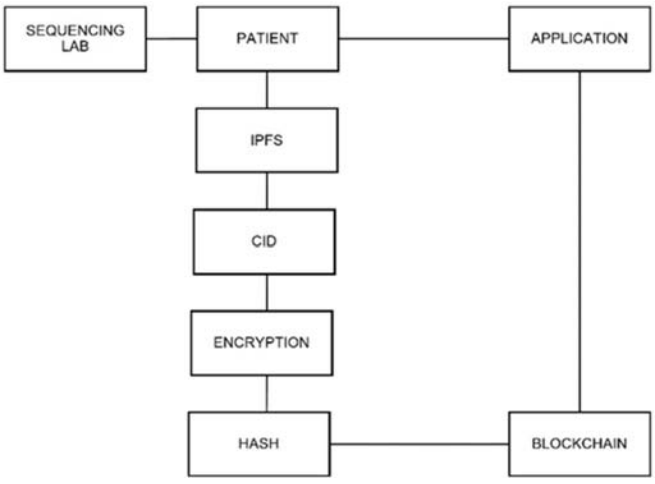

Fig. 2 Pathway for transfer of genomic files

B. Uploading

With the 'upload file' feature, users can choose to upload their entire genomic file or selectively grant access to specific sections. They can submit their queries for immediate

processing or save them for future use.

C. Storage

Users can either use their files immediately or store them in the software for later access. Files can be saved on the app's servers for easy access without encryption, automatically deleting after three days. Premium members get NucleoLocker for secure genomic data storage, plus advanced features like blockchain data retrieval and enhanced security. They can also process genomic data stored elsewhere securely. To reduce file size, a data compression algorithm can be used, expanding files when needed.

D. Processing

Users often seek specific analyses to understand their physical traits or predict future health risks. For example, someone might want to know why their eyes are a certain color or what their genetic predisposition is for certain diseases like diabetes, especially if they lead a sedentary lifestyle. To make this process easier, a data abstraction method hides complex operations, offering a user-friendly experience. A Graphical User Interface (GUI) visually shows how things work behind the scenes, keeping users engaged.

The processing involves analyzing user queries, starting with expanding compressed genomic files. Recurrent neural networks, like Long Short-Term Memory (LSTM), excel at handling sequential data, which in this case involves genetic information.

Advanced AI techniques such as 'Fuzzy Logic' and 'Expert Systems' enhance this analysis. Expert systems mimic human decision-making in specific fields, using extensive knowledge of human DNA and genetic factors to answer user queries and predict future health issues. Fuzzy logic assigns degrees of certainty to conditions or predictions, indicating the severity of current or potential health issues.

In order to derive insights based on an individual's genome, an integrated framework combining an expanding and abbreviating algorithm, LSTM, expert system, and fuzzy logic may be constructed.

An example code is given to help understand how the application processes information. This code is straight-forward and meant for understanding purposes; however, the actual implementation might vary significantly.

```
from Bio import SeqIO
from Bio.Seq import Seq
from Bio.SeqRecord import SeqRecord

# Step 1: Retrieve user's genome data
user_genome_file = "user_genome.fasta" # Replace with
actual file path or input method
user_genome = SeqIO.read(user_genome_file, "fasta")

# Step 2: Retrieve reference genome
reference_genome_file = "reference_genome.fasta" #
Replace with actual file path or download from a database
reference_genome = SeqIO.read(reference_genome_file,
"fasta")

# Step 3: Compare genomes
def compare_genomes(user_genome, reference_genome):
    mismatches = []
    for i, (user_base, ref_base) in
enumerate(zip(user_genome, reference_genome)):
        if user_base != ref_base:
            mismatches.append((i, ref_base, user_base))
    return mismatches
mismatches = compare_genomes(user_genome.seq,
reference_genome.seq)

# Step 4: Analyze variations
for pos, ref_base, user_base in mismatches:
    print(f"Mismatch at position {pos}: Reference base is
{ref_base}, User base is {user_base}")
    # Perform further analysis using databases like
dbSNP, ClinVar, OMIM, GWAS Catalog, HapMap, etc.

# Step 5: Display results
# Share the results of the analysis with the user
```

The code provided hereafter is a simplified illustration of application. However, integrating with databases like dbSNP, ClinVar, OMIM, etc., would require additional code and potentially API requests to retrieve relevant information.

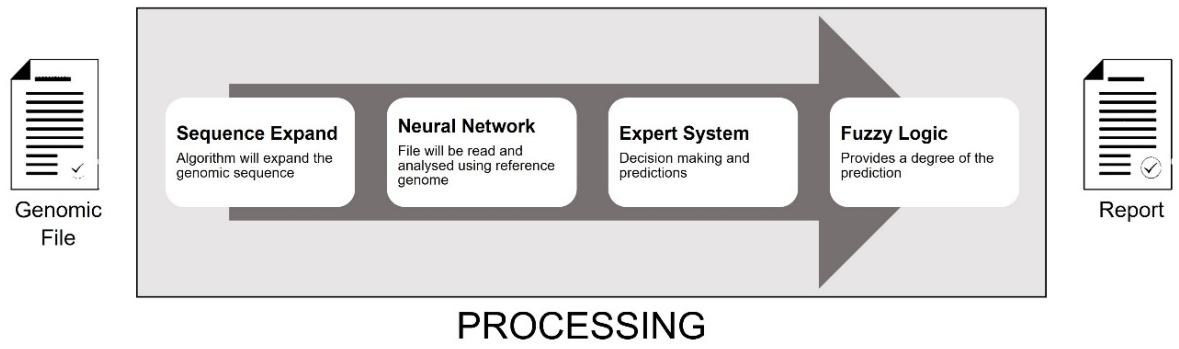

Fig. 3 Steps involved in transforming a genomic file into a report

E. Insights and Recommendation

This system would use data to give personalized recommendations using expert systems and fuzzy logic. It will recommend changes in diet and exercise, more tailored to each person's needs for improved health. User would get detailed reports with graphs to understand and take necessary steps, aided by built-in data visualization tools.

F. Download and Delete

Once the report is generated, the user could choose to download it to their device or delete it from the program if they wish. Moreover, the report would be password protected to enhance security.

G. Save

This functionality offers multiple ways to save reports within the application:

- Easy Access: Reports can be stored without encryption if kept on the server.
- NucleoLocker: A "NucleoLocker" option within the application could provide a secure locker for premium users to store their genomic data.
- Blockchain: Users could choose to save their genomic files, along with insights and recommendations, in an encrypted form on the blockchain. The data are encrypted, and a generated hash is uploaded to the blockchain, ensuring tamper-proof storage and immutability.

H. Sharing

This feature allows users to securely share medical reports with doctors. Using Blockchain technology, reports are uploaded and encrypted for safe sharing. Users can grant access to their reports using a combination of their public key and the doctor's private key. This method ensures data security and integrity through several steps. Alternatively, users can submit a request form within the app, specifying the recipient, time, and amount of data. The report is encrypted and sent to the recipient within the specified timeframe. To decode the report, the recipient needs a one-time password provided by the user.

I. Advanced Antenatal Prognosis

This program also offers prenatal diagnosis using the genetic information of the biological parents. It predicts general details about the child's genetic makeup and final physical traits, focusing on the chances of diseases or disabilities. It uses Regression training, a machine learning method, to understand patterns in datasets that include target variables and input features. This helps identify important factors like single nucleotide changes (SNPs), genetic differences, inheritance patterns, and markers. To produce forecasts of a high accuracy, however, an exceptionally high caliber and amount of data will be required.

J. Testing, Evaluation and Validation

A set of data, including clinical information and training sessions, is used to teach this model. Its effectiveness can be assessed by looking at metrics like accuracy, precision, recall, F1-score, and ROC-AUC curve. These metrics can help in finding weaknesses in the model. It is important to test the model's accuracy and precision in predicting illnesses or answering questions. This helps in determining a threshold to classify a specific condition as positive or negative.

K. Feedback Engine

Users can share their opinions based on insights they receive. This helps verify the accuracy of recommendations and insights they get. Users can also provide details about their environment, improving the accuracy of insights generated. Users can give feedback on the application's usability, complexity, GUI, and more within the system.

L. Regular System

This system is crucial for the application because our knowledge of the genome and its relationship with various factors is constantly changing. To stay relevant, the system needs regular updates based on user feedback and new information. This ensures that the application can continuously provide accurate insights and recommendations.

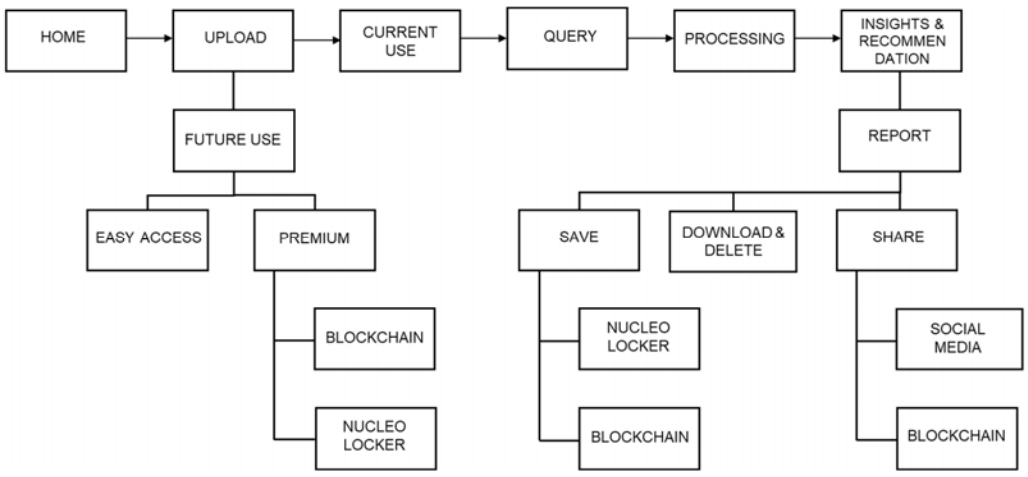

Fig. 4 Application workflow, surface-level components, and involvement of blockchain

IV. Data Storage

The HGP officially wrapped up in 2003, but its full potential remains untapped even after more than two decades. One major hurdle is the sheer volume of data humans produce which requires storage. The internet is bursting with vast amounts of information, far beyond our current storage capabilities. This issue is especially pronounced in genome projects. As highlighted by a blog on 3billion.io, an astounding 40 exabytes

of genomic data is generated every year—40 times more than the data created by YouTube annually [6]. To address this massive storage challenge, traditional methods like cloud storage, SSDs, and hard disks are falling short. These systems struggle with the exponential data growth and have inherent limits. DNA-enabled data storage, a groundbreaking technology that promises a solution to the problem [3], [4]. This method involves creating synthetic DNA from nucleotide sequences derived from binary data. Essentially, the binary data (comprising 0s and 1s) of a digital file are encoded into the four nucleotide bases—A, T, G, and C—using a predefined encoding scheme. For example, we consider this encoding scheme:

A: 01
T: 00
G: 10
C: 11

With this system, the binary data of an image file can be transformed into a sequence of A, T, G, and C. This sequence can then be used to create synthetic DNA, which can be stored for extended periods with minimal energy consumption under optimal conditions. When needed, this artificial DNA can be decoded back into its original binary form using the same encoding technique. This approach not only promises to revolutionize data storage but also ensures that our ever-growing digital information can be preserved efficiently for the future.

A. Limitations with DNA Enabled Data Storage

The 'storage issue' may be substantially resolved and by 2040, it may be considered that one kilogram of DNA could store all the data in the world, estimated to be around 3 x 10^{24} bits [4]. However, the major drawbacks associated with this technology cannot be overlooked. Among them are:

1. Limited Access

This cutting-edge technology requires a complex setup, involving computer algorithms to encode and decode binary data (0s and 1s) into the nucleotides A, T, G, and C, along with DNA synthesizers to produce synthetic DNA. Ensuring equitable distribution and universal availability of such sophisticated equipment and protocols is a daunting task. Additionally, sharing data becomes problematic as this method necessitates the physical storage of data, making quick information exchange over long distances difficult.

2. Data Retrieval

For data stored in artificial DNA (a-DNA) to be used, it must be decoded back into binary form. This extra step can be tedious and presents a significant drawback to this approach.

3. Data Misuse

Unlike digital data, which are stored on hard drives or in data centers and are accessible anytime, anywhere, a-DNA data are physically stored. This creates a risk where the storage authority might exploit the data owner, limiting their control over their own information.

4. Cost Constraints

Storing information in DNA requires advanced instruments and technologies, making it more expensive than traditional hard drives. Despite its potential to solve storage challenges, the high costs and technical glitches make this technology impractical for widespread use. Until new advancements can overcome these limitations, digital storage will remain the standard, preserving essential features like data retrieval, sharing, safety, security, speed, and efficiency.

5. Data Loss

Negligence in proper storage might cause DNA degradation, leading to potential data loss, although this is a rare occurrence.

B. Data Abbreviation

To tackle the existing technical glitches and potentially reduce the size of genomic sequencing data, a method has been devised. Genomic data can be stored in two primary formats: Fasta files and 2-bit files. Fasta files, commonly found as text files or flat files, can be opened with nearly any text editor. The FASTA format, named after the software that quickly searches large amounts of sequence data, uses 8 bits or 1 byte per nucleotide, making it the most popular format for storing sequences.

A typical human genome, with 3 billion base pairs, would require 3 gigabytes (GB) of storage in the FASTA format, as each base pair occupies one byte. When considering the entire human population of 8 billion, this amounts to an astronomical 24 billion GB, not including additional data like formatting, annotations, and metadata.

An alternative storage method is the 2-bit file format, a binary format that encodes DNA nucleotides using only two bits per base pair, significantly more efficient than the text-based FASTA format. For the human genome, this equates to 0.75 GB (750 megabytes) per individual, and for the global population, a total of 12 billion GB.

To further address the storage challenge, 'Data Abbreviation' has been proposed. This method involves compressing repetitive sequences using mathematical representations, thereby managing and storing vast amounts of genomic data more efficiently. Given that less than 2% of DNA codes for proteins and the remaining 98% consist of repetitive sequences, abbreviating these repeats can significantly reduce file sizes [7]. For example, a sequence with 4267 consecutive 'A' nucleotides can be abbreviated as A[4K;2C;6X]AAAAAAA, where 'K' represents 1000, 'C' represents 100, and 'X' represents 50.

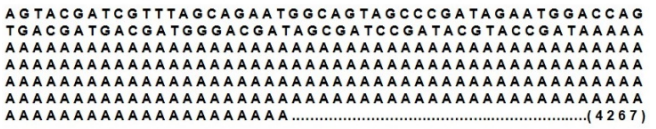

Fig. 5 An arbitrary DNA sequence containing repeating nucleotides is presented for illustration

By employing such innovative techniques, the technical

barriers of genomic data storage can be overcome, making individual genome sequencing more feasible and efficient:

```
def shorten_dna_string(dna_string):
    shortened_string = ""
    current_base = dna_string[0]
    count = 1

    for base in dna_string[1:]:
    if base == current_base:
    count += 1
    else:
    shortened_string += f"{current_base}("
    shortened_string += "(" + (count // 1000000) * "M" +
")" if count >= 1000000 else ""
    count %= 1000000
    shortened_string += "(" + (count // 1000) * "K" + ")"
if count >= 1000 else ""
    count %= 1000
    shortened_string += "(" + (count // 500) * "D" + ")"
if count >= 500 else ""
    count %= 500
    shortened_string += "(" + (count // 100) * "C" + ")"
if count >= 100 else ""
    count %= 100
    shortened_string += "(" + (count // 50) * "L" + ")" if
count >= 50 else ""
    count %= 50
    shortened_string += "(" + (count // 10) * "X" + ")" if
count >= 10 else ""
    count %= 10
    shortened_string += "(" + (count // 5) * "V" + ")" if
count >= 5 else ""
    count %= 5
    shortened_string += ")" * count
    current_base = base
    count = 1

    shortened_string += f"{current_base}("
    shortened_string += "(" + (count // 1000000) * "M" +
")" if count >= 1000000 else ""
    count %= 1000000
    shortened_string += "(" + (count // 1000) * "K" + ")"
if count >= 1000 else ""
    count %= 1000
    shortened_string += "(" + (count // 500) * "D" + ")"
if count >= 500 else ""
    count %= 500
    shortened_string += "(" + (count // 100) * "C" + ")"
if count >= 100 else ""
    count %= 100
    shortened_string += "(" + (count // 50) * "L" + ")" if
count >= 50 else ""
    count %= 50
    shortened_string += "(" + (count // 10) * "X" + ")" if
count >= 10 else ""
    count %= 10
    shortened_string += "(" + (count // 5) * "V" + ")" if
count >= 5 else ""
    count %= 5
    shortened_string += ")" * count

    return shortened_string

# Example usage
dna_string = "AAAATTTTGGGGCCCAAAATTTT……"
shortened_string = shorten_dna_string(dna_string)
print(shortened_string)
```

Thus, the resulting sequence is formed as follows:

AGTACGATCGTTTAGCAGAATGGCAGTAGCCCGATA
GAATGGACCAGTGACGATGACGATGGGACGATAGCG
ATCCGATACGTACCGAT A[4K;2C;6X] AAAAAAA

Fig. 6 The genomic data file before and after Data abbreviation

C. Storage of Genomic Data on Blockchain

Next to personal genome sequencing and storage, security stands as a towering challenge. The allure of personal genomic data as a honeypot for data hackers is undeniable, given its potential for substantial financial gain. Therefore, safeguarding these data during storage and transmission becomes critical. One promising solution lies in blockchain technology, revolutionizing how genomic information is utilized and protected.

By merging blockchain with genomics and healthcare, patients and researchers gain greater empowerment. Blockchain's decentralized nature ensures that genomic data are stored securely across numerous network nodes, maintaining its integrity and accessibility. This transformative technology not only fortifies data against tampering but also facilitates rapid, secure global transmission of genomic files. This promises to amplify collaboration in genomic research and empower healthcare providers worldwide to deliver tailored treatments and precision medicine.

V. Discussion

Attempts have been made to use the reference genome from the HGP for various genomic analyses. The use of AI, ML, and blockchain in genomics is not new; for instance, neural networks are applied in cancer research, genomics, and protein analysis [8]. These efforts have been piecemeal until now, with a comprehensive framework developed using AI and ML for personalized analysis of the entire human genome. This tailored approach offers deeper insights into individual genetic differences compared to general health advice.

Efforts to reduce the size of genomic files have focused on comparing and compressing similar DNA sequences, given that 99.95% of DNA is identical across individuals [9], [10]. In this study, a method has been devised to compress whole genomes efficiently, reducing processing time and eliminating the need for a reference genome. This straightforward approach can be

applied multiple times across different stages of genetic data processing, not just for storage but also for various analytical purposes.

Blockchain technology is utilized in genome sequencing to enhance data security and privacy during transfers. This research proposes using blockchain alongside smart contracts to give genomic data owners control over data access and transfer. Integrating homomorphic computation with secure protocols allows users to securely query and discover relevant data [11]. In healthcare, blockchain optimizes processes, lowers costs, improves patient outcomes, and ensures secure data sharing [12]. This study suggests storing entire genomes securely using file systems like IPFS and encryption, ensuring data integrity and user ownership through cryptographic methods.

VI. Conclusion

Many researchers have explored different technologies for storing genomic data. However, current DNA data storage methods have limitations, especially when data need to be shared across distant locations and organizations. This has led to a need for digital-only storage solutions, despite the large volume of genomic data, which is difficult to manage.

This study proposes practical solutions to these challenges by focusing on compressing and reducing the size of genomic data files. Techniques for summarizing data help to save space by condensing genomic information, making it easier to store and use in applications that focus on genomes. Algorithms developed in this study efficiently manage the length of genetic sequences, improving storage efficiency.

The study also presents the 'Your Genome App', a tool aimed at advancing healthcare through personalized genomics. This app allows individuals to gain valuable insights from their genetic information, predicting potential health issues early. By providing tailored recommendations based on genetic profiles, the app aims to reduce reliance on medication or surgery, enhancing preventive healthcare [13].

To ensure data security and integrity, the app integrates blockchain technology. Blockchain's decentralized structure protects genomic data during transport and storage, enabling secure data exchange while maintaining privacy and ensuring data authenticity.

While the convergence of genomics and digital technology, especially in trending topics like AI, is promising, practical implications and regulatory challenges remain unclear. Ethical concerns also arise due to the sensitive nature of genomic data, highlighting the importance of robust security measures like blockchain to safeguard against potential breaches.

References

[1] Amaral, P., Carbonell-Sala, S., De La Vega, F.M., Faial, T., Frankish, A., Gingeras, T., Guigo, R., Harrow, J.L., Hatzigeorgiou, A.G., Johnson, R., Murphy, T.D., Pertea, M., Pruitt, K.D., Pujar, S., Takahashi, H., Ulitsky, I., Varabyou, A., Wells, C.A., Yandell, M., Carninci, P., & Salzberg, S.L. (2023). The status of the human gene catalogue. Nature, 622(7981), ArXiv. Online Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10055485/#R2

[2] Gencode. (2024). Gencode - Human Release 45 Statistics. Online. Available at: https://www.gencodegenes.org/human/stats_45.html.

[3] Goldman, N., Bertone, P., Chen, S., Dessimoz, C., LeProust, E.M., Sipos, B., & Birney, E. (2013). Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature, 494(7435), 77–80. https://doi.org/10.1038/nature11875.

[4] Panda, D., Molla, K.A., Baig, M.J., Swain, A., Behera, D., & Dash, M. (2018). DNA as a digital information storage device: hope or hype? 3 Biotech, 8(5). https://doi.org/10.1007/s13205-018-1246-7.

[5] Aganezov, S., Yan, S.M., Soto, D.C., Kirsche, M., Zarate, S., Avdeyev, P., & Koren, S. (2022). A complete reference genome improves analysis of human genetic variation. Science, 376(6588). https://doi.org/10.1126/science.abl3533.

[6] 3billion.io. (2022). Genetic testing for rare diseases company | 3billion. Online. Available at: https://3billion.io/blog/big-data-among-big-data-genome-data.

[7] Buehler, J. (2024). The Complex Truth About 'Junk DNA'. Quanta Magazine. Online. Available at: https://www.quantamagazine.org/the-complex-truth-about-junk-dna-20210901/.

[8] Caudai, C., Galizia, A., Geraci, F., Le Pera, L., Morea, V., Salerno, E., & Colombo, T. (2021). AI applications in functional genomics. Computational and Structural Biotechnology Journal, 19, 5762–5790. https://doi.org/10.1016/j.csbj.2021.10.009.

[9] Deorowicz, S., Danek, A., & Niemiec, M. (2015). GDC 2: Compression of large collections of genomes. Scientific Reports, 5(1). https://doi.org/10.1038/srep11565.

[10] Wang, C., & Zhang, D. (2011). A novel compression tool for efficient storage of genome resequencing data. Nucleic Acids Research, 39(7), e45. https://doi.org/10.1093/nar/gkr009.

[11] Yakubu, A.M., & Chen, Y.P.P. (2022). A blockchain-based application for genomic access and variant discovery using smart contracts and homomorphic encryption. Future Generation Computer Systems, 137, 234-247. https://www.sciencedirect.com/science/article/abs/pii/S0167739X22002400#:~:text=We%20use%20blockchain%20transactions%20and,securely%20discover%20DOs%20of%20interest.

[12] Mackey, T.K., Kuo, T.T., Gummadi, B., Clauson, K.A., Church, G., Grishin, D., Obbad, K., Barkovich, R., & Palombini, M. (2019). 'Fit-for-purpose?' – Challenges and Opportunities for applications of blockchain technology in the future of Healthcare. BMC Medicine, BioMed Central 17, 68. https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1296-7.

[13] Garg, H.K., & Vyas, A. (2024). Exploring Genome Projects and Beyond - Untwining the Past, Present & Future for Transformative Applications. Online. Available at: https://zenodo.org/records/10466512.

Anuja Vyas, a sophomore at the Institute for Excellence in Higher Education, Bhopal, India, is on a quest in Biotechnology. While working on an enthralling research project, 'Codes beyond Bits & Bytes - A Blueprint of Artificial Life', she coined the term 'Genetronics'. Anuja concocted an innovative application that impeccably detects variations within a genome compared to the standard one.

Aikel Indurkhya is an accomplished undergraduate student earning a Bachelor's degree in Biotechnology at the Institute for Excellence in Higher Education, Bhopal, India. She received the prestigious Director's Medal (2024) for her outstanding achievements in academics and extracurricular activities. Aside from her rigorous academic initiatives, Aikel has previously worked on projects focusing on blockchain technology, encryption, storage security and transaction methodologies.

Dr. Hari Krishna Garg is a Professor at Department of Biotechnology, Institute for Excellence in Higher Education, Bhopal, India. He has 31 years of experience in Research and Academics.

He has authored Books on Research Techniques in Molecular Biology, Genetics & Biotechnology, Avian Cytogenetics, Genetic Engineering, and Instrumentation for European & US scholars. He has attended >60 Seminars and authored more than 100 Research Papers.