

SIMULATING NEXT-GENERATION DATA STORAGE ARCHITECTURES USING DNA, GRAPHENE, AND NEURAL ENCODING

Hamza Arif ¹, Muhammad Tanveer Meeran ², Sabiha Anum ³, Assad Latif ⁴

¹ School of Business and Economics (SBE), University of Management and Technology (UMT), Lahore, Pakistan.

² Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu, Malaysia.

³ Administrative sciences (Business Management), Boston University (Metropolitan College)

⁴ School of Management and Economics, North China University of Water Resources and Electric Power, Zhengzhou 450046, China

DOI: <https://doi.org/>

Article History

Received on 20 Nov 2025

Accepted on 20 Dec 2025

Published on 18 Jan 2026

Copyright @Author

Corresponding Author: *

Tanveer_meeran@yahoo.com

Abstract

In today's data-driven world, the rapid growth of digital information is placing increasing pressure on traditional storage technologies, which struggle to meet demands for scalability, energy efficiency, and long-term sustainability. To address these emerging challenges, this paper introduces Neuron, a cross-platform simulation environment developed to examine future-oriented data encoding concepts such as DNA-based storage, graphene-inspired atomic memory, and neural signal-driven interfaces. Instead of physically storing data in these advanced media, Neuron focuses on simulating their encoding mechanisms through interactive animations, real-time analytical feedback, and immersive 3D visual representations. The platform is built using contemporary web and desktop frameworks, including React.js, Tauri.js, and Supabase, allowing it to effectively connect theoretical research with hands-on educational exploration.

Within Neuron, users can upload digital files, initiate encoding simulations, and observe how data may be transformed into DNA nucleotide patterns, graphene-like grid formations, or neural signal pathways. The system incorporates features such as secure file handling, role-based user access, and real-time dashboards, making it suitable for both independent learners and academic or research institutions.

This paper presents the overall system architecture, simulation techniques, and design principles behind Neuron, while

emphasizing its role in promoting awareness and understanding of sustainable and futuristic data storage paradigms. By offering simulation rather than physical implementation, Neuron provides an accessible and interactive means to explore how digital information could one day be represented at molecular or neural levels. Index Terms—DNA Encoding, Graphene Memory, Neural Data Simulation, Brain-Computer Interface, Cloud Simulation, Tauri.js, Supabase, Real-Time Visualization, Role-Based Access Control

INTRODUCTION

The exponential rise in digital data across scientific, industrial, and personal domains has exposed critical limitations in traditional data storage systems, particularly concerning scalability, sustainability, and long-term reliability. Magnetic and silicon-based mediums such as hard disk drives (HDDs), solid-state drives (SSDs), and cloud-based solutions are rapidly approaching physical and technological constraints. As these limitations become more evident, the exploration of next-generation storage technologies such as DNA-based memory [1]–[6], graphene-based memory systems [7]–[12], and brain-computer interface-driven neural storage [13]–[20] has emerged as an essential research frontier.

Despite rapid scientific advancements, the real-world application of these emerging storage formats remains hindered by technical complexity, cost, and lack of user-friendly experimentation environments. Most existing research in this field remains confined to laboratory simulations or abstract theoretical models. The absence of accessible simulation tools prevents wider participation in understanding and developing futuristic storage technologies. In response to this gap, we introduce Neuron, a cloud-connected, cross-platform desktop application that simulates the encoding of binary data into DNA sequences, graphene atomic structures, and neural signal patterns.

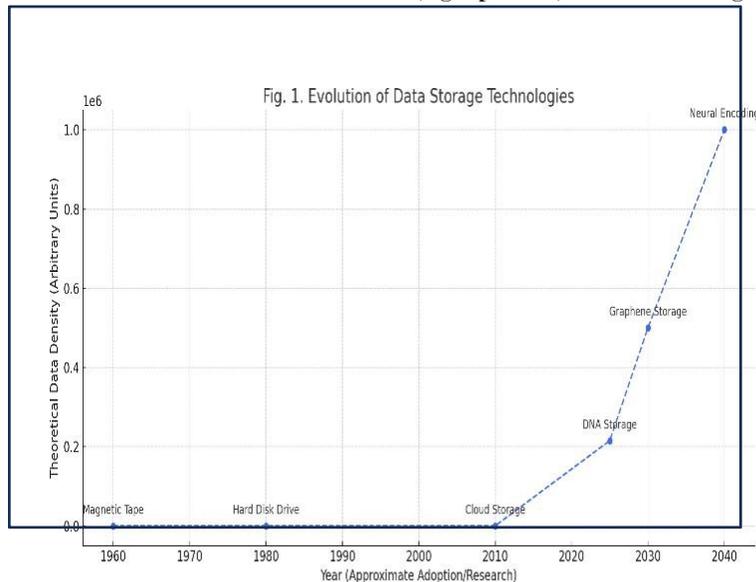
Neuron aims to transform how users interact with the concept of data storage by providing real-time 3D visualization, conversion simulation, and analytics dashboards. Built using modern frameworks such as React.js, Next.js, Supabase, and Tauri.js, the platform not only supports conventional file uploading but also enables users to simulate advanced storage workflows through an immersive, animated interface. Conversion processes are purely virtual, with no physical data writing to DNA, graphene, or neural substrates, preserving safety while enabling education and experimentation.

Preprocessing techniques ensure a streamlined user experience, and modular architecture supports performance across Windows, macOS, and Linux. The platform also features role-based access control, allowing differentiated user experiences for Admins and general Users. While Neuron simulates theoretical storage paradigms, its design is grounded in real academic literature and technical feasibility studies. The primary limitations encountered during development include handling complex animations under limited hardware conditions, platform-specific deployment issues, and managing real-time state transitions across conversion stages.

Nonetheless, Neuron presents a robust proof-of-concept simulation environment that supports research, education, and early prototyping for futuristic storage systems. It paves the way for better understanding the

convergence of biotechnology, material science, and neuro-computing in data handling. By simulating real encoding processes for DNA, graphene, and neural formats, it bridges the gap between theoretical research and practical visualization, making next-generation storage concepts more accessible and demonstrable.

Fig. 1. shows the evolution of data storage technologies from traditional magnetic systems to advanced molecular and neural encoding methods. Neuron simulates DNA, graphene, and brain-signal-based formats to explore



future storage paradigms.

Figure 1 Data Storage Evolution

The remainder of this paper is structured as follows: Section II presents a literature review of related work in DNA, graphene, and neural-based storage models. Section III discusses the methodology and simulation workflow employed in the Neuron platform. Section IV outlines system design and implementation, while Section V presents results and evaluations. Section VI addresses current limitations and challenges. Finally, Section VII concludes the study and proposes directions for future enhancement.

I. Literature Review

As global data generation continues to grow at an exponential rate, fueled by the expansion of cloud computing, artificial intelligence, IoT devices, and high-resolution multimedia, conventional digital storage systems are approaching their physical and technical limits. Current storage infrastructures such as hard disk drives (HDDs), solid-state drives (SSDs), and cloud-based servers are increasingly challenged by the demand for higher capacity, faster access speeds, lower energy consumption, and improved long-term durability. As these limitations become more pronounced, researchers and technologists are urgently exploring unconventional and disruptive alternatives that can overcome the shortcomings of traditional storage media.

Among the most promising next-generation storage paradigms are DNA-based memory systems, graphene-enabled atomic storage, and neural encoding mechanisms inspired by brain-computer interfaces (BCIs). These approaches aim to leverage the unique characteristics of biological molecules, nanoscale materials, and neural signal patterns to drastically improve data density, resilience, and sustainability. DNA offers unparalleled information density and long-term archival potential; graphene, with its two-dimensional carbon lattice, presents opportunities for atomic-level data encoding and energy-efficient memory; while neural-based storage explores

how binary data might one day be represented using cognitive signal structures or spike trains.

Although these technologies remain in early stages of development and are not yet commercially viable, their conceptual frameworks are rapidly evolving through experimental research and simulation environments. These innovations have not only sparked interdisciplinary collaboration among fields such as synthetic biology, materials science, and neuroscience, but have also opened new paradigms in how we understand and interact with the future of digital memory. Simulation platforms like Neuron emerge in this context not as replacements for physical encoding systems but as essential educational and exploratory tools that allow researchers, students, and technologists to interact with, visualize, and prototype these cutting-edge storage concepts in a virtual setting.

A. DNA-Based Storage

DNA, the biological molecule that encodes genetic information in living organisms, has been identified as a potential storage medium due to its unparalleled data density and stability. One gram of synthetic DNA can theoretically hold up to 215 petabytes of data [1]. Church et al. [2] were among the first to demonstrate encoding digital data into DNA sequences, translating binary information into nucleotide bases. Later, Goldman et al. [4] improved this method by introducing redundancy and efficient error correction. Erlich and Zielinski's DNA Fountain technique pushed the field further by achieving robust and scalable data retrieval with minimal errors [1]. Despite its promise, DNA-based storage faces practical limitations related to high costs of synthesis, error-prone sequencing, and slow write/read speeds.

B. Graphene and Atomic-Scale Storage

Graphene, a two-dimensional material composed of a single layer of carbon atoms, is another frontier in advanced storage. Its atomic thinness, high conductivity, and mechanical strength make it a promising candidate for ultra-dense memory systems. Graphene has been studied as a material for resistive RAM (RRAM) and magnetic storage coatings, showing up to 10^8 higher capacity than traditional materials [8], [10]. Dwivedi et al. explored graphene's potential in increasing storage density using nanocoating, enabling data storage at near-atomic scales [7]. While commercial adoption is limited by fabrication challenges, the scientific community continues to explore its use in scalable, high-speed, energy-efficient storage [9], [11], [18].

C. Neural Encoding and Brain-Computer Interfaces

The third paradigm investigated in Neuron is neural or brain signal-based encoding. Brain-computer interfaces (BCIs) are advancing rapidly in areas such as prosthetics, communication devices, and neurofeedback. Researchers like Chakraborty and Bhunia have theorized methods of encoding binary data using spike-train neural patterns [13], enabling elastic, signal-driven memory systems. Simultaneously, projects like Neuralink have demonstrated invasive and non-invasive BCIs that allow for high-bandwidth transmission between brain and machine [15], [19]. Although real-time brain-based data storage is still speculative, the concept is supported by increasing research into EEG/MEG signal processing, cognitive informatics, and memory modeling [16], [17].

D. Gaps in Practical Implementation

Despite the theoretical and experimental progress in these domains, existing systems remain difficult to access for hands-on learning, testing, or prototyping. There is a lack of simulation tools that abstract the complexity of physical encoding while preserving the conceptual integrity of these advanced storage paradigms. Neuron

addresses this gap by offering an educational and experimental platform that simulates data transformation into these formats, thereby democratizing access to futuristic data handling techniques.

II. Methodology

In the context of research, methodology outlines the step-by-step procedures, techniques, and tools employed to collect data, analyze information, and draw conclusions (Figure 2). Below is a step-by-step process for implementing the suggested method. Additionally, each stage is briefly discussed.

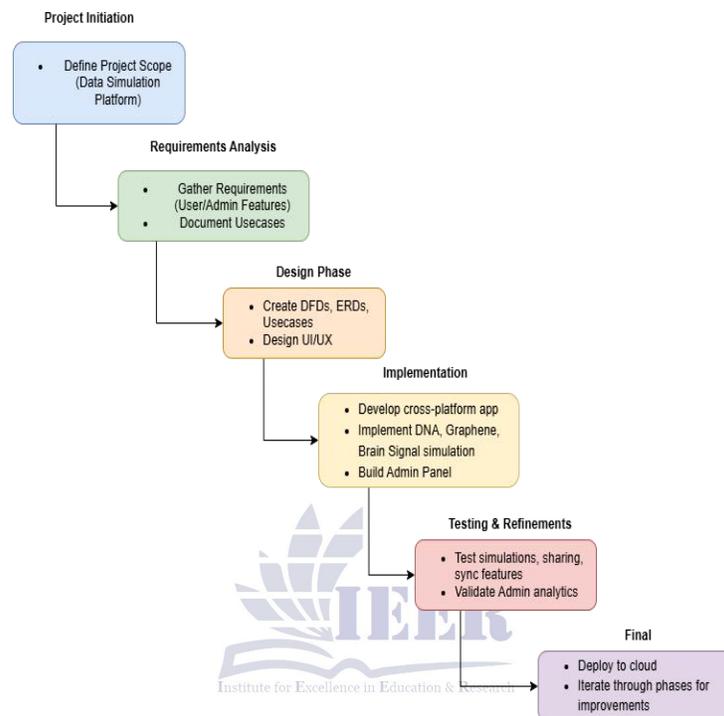


Figure 2 Proposed Methodology

A. System Architecture Overview

Neuron is built using a hybrid technology stack comprising Next.js for frontend logic and routing, React.js for UI components, Tauri.js for secure desktop packaging, and Supabase for real-time backend services including authentication, file storage, and database management. The application architecture is modular, with clearly separated layers for:

- **User Interface (UI):** Built with React, Tailwind CSS, and Framer Motion for smooth animations and transitions.
- **State Management:** Implemented using Zustand to manage complex real-time states during simulations.
- **File Handling:** Users upload files which are stored securely in Supabase's storage bucket and metadata is synced to the database.
- **Role Management:** Admins and users are separated through authentication guards and role-based UI rendering.

B. Data Simulation Workflow

Neuron simulates the transformation of binary files into three advanced formats: DNA sequences, graphene atomic grids, and brain signal patterns. Unlike traditional UI-only simulations, Neuron integrates actual format-specific encoding logic that processes binary input into symbolic storage representations based on principles found in DNA-based memory systems, resistive nanomaterials, and neural signal modeling.

1) DNA Storage Simulation

Upon file upload, the file is parsed and converted into binary data, which is then translated into synthetic

DNA letter	Binary Representation
A	00
C	01
G	10
T	11

nucleotide sequences using a basic 2-bit encoding scheme mention in Figure 3.

Figure 3 DNA Encoding

This mapping mimics encoding techniques such as those used in DNA Fountain [1]. The resulting sequence is visualized with animated helices and base-pair highlights. Though the output is not physically synthesized, the system does generate a valid DNA-like string using real binary data, offering users an interactive approximation of how information might be stored in synthetic DNA strands.

2) Graphene Storage Simulation

Graphene In this simulation, Neuron uses a resistance-based encoding scheme to simulate binary information as high- and low-resistance nodes. The system first extracts a binary sample from the file, and each bit is encoded as either "high" or "low":

- 1 → high
- 0 → low

This behavior emulates the switching mechanism used in graphene-based RRAM devices, where the resistance level of a memory cell indicates a binary state. While the simulation does not operate on actual graphene, it provides a conceptually accurate representation of how data might be encoded at the atomic scale in future devices.

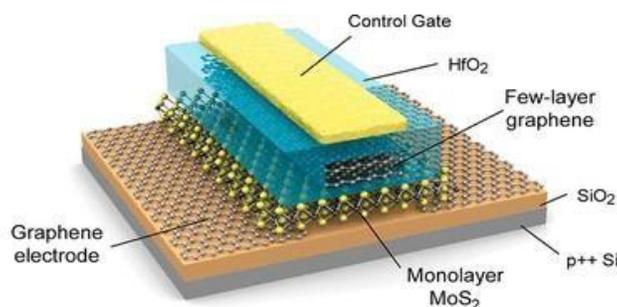


Figure 4 Schematic of a graphene-based memory cell structure

3) Neural Encoding Simulation

For neural signal encoding, Neuron processes image, video, audio, or PDF files into intermediate data forms such as pixel arrays, audio buffers, or byte streams. It then synthesizes EEG-like waveform signals representing these inputs, formatted as JSON objects that include:

- File metadata (name, size, type)
- Timestamp and signal origin
- Encoded signal data (in base64)
- Simulated signal confidence score

This mimics spike train modeling and synthetic EEG signal generation found in BCI research [13], [16], providing an engaging approximation of how brain signals could represent digital information.

C. User Interaction and Conversion Flow

Each conversion flow follows a consistent structure:

1. *Upload Phase*: File type is validated and uploaded to Supabase.
2. *Selection Phase*: User selects one of the three simulation types.
3. *Animation Phase*: Progress bar, staged icons, and animated background render the transformation in real-time.
4. *Completion Phase*: A success message with confetti and a checkmark Lottie file indicates completion. Optionally, the user can download a pseudo-format output (not real data).
5. *Role Control*: Admins can view file analytics, user stats, and shared file activity through the Admin Panel.

D. Performance Optimization

To support smooth rendering on desktop environments, optimizations include:

- Lazy loading for animations
- Controlled frame rates using `requestAnimationFrame`
- Memory-efficient storage management via Supabase buckets

These enhancements improve cross-platform consistency and minimize lags in animations.

III. System Design and Implementation

Neuron is designed as a modular, cross-platform desktop application with cloud capabilities. Its system architecture separates frontend simulation, backend data management, and role-based access control into distinct layers. This design ensures that simulation rendering, user authentication, and file management remain decoupled and maintainable, improving scalability and security.

A. Frontend Design and Technologies

The user interface is built with Next.js and React.js, offering a fast, responsive, and scalable architecture. Styling and layout components use Tailwind CSS, while Framer Motion and Lottie animations handle micro-interactions and immersive visuals. The main UI components include:

- **Dashboard**: Centralized interface where users can upload files, monitor usage, and choose storage formats.
- **Conversion Modal**: A modal component that visually guides users through conversion stages with animated feedback.
- **Admin Panel**: A secure section for monitoring users, file logs, and system analytics.

Routing is handled client-side using Next.js dynamic routes. State management uses Zustand, chosen for its lightweight and reactive state updates, particularly suited for simulation rendering.

B. Backend and Database Logic

The backend relies on Supabase, an open-source Firebase alternative offering:

- **Authentication:** JWT-based user login and role assignment (admin/user).
- **Storage:** File uploads and storage buckets with signed URLs.
- **Database:** PostgreSQL database for managing file metadata, user roles, and order records.

Supabase's real-time capabilities are used to reflect file upload status and user activity across devices. Backend operations like fetching order history or managing shared files are abstracted via Supabase client SDKs, securely integrated into frontend components.

C. File Conversion and Simulation Modules

Once a file is uploaded, users can initiate a simulation by selecting one of the three available storage formats. Each format is powered by a dedicated rendering module:

- **DNA Module:** Uses base-encoded string mapping with animated helix rendering and pair-matching.
- **Graphene Module:** Displays animated 2D atomic grids where data bits toggle simulated carbon nodes.
- **Neural Module:** Converts binary streams into waveform patterns rendered using animation.

These modules are organized as reusable React components, allowing smooth transitions between simulations while sharing common utility functions like binary conversion, loader progress, and success indicators.

D. Role-Based Access Control

Neuron incorporates a simple but effective access control system:

- **Admin Users:** Can access user lists, order statistics, file share logs, and dashboard analytics.
- **General Users:** Can upload files, run simulations, and share data with other users.

Access is managed using conditional rendering in React, based on the user's Supabase role. Unauthorized users are redirected automatically using middleware guards and authentication hooks.

E. Desktop Deployment with Tauri.js

The platform is packaged into a native desktop application using Tauri.js, which provides secure system integration and lightweight binaries. Compared to Electron, Tauri offers a smaller footprint and better performance, especially on low-spec machines. Tauri allows:

- File system interaction (sandboxed)
- Local storage
- Cross-platform builds (Windows, macOS, Linux) Tauri's system tray integration and auto-updates support are also planned in future versions of Neuron.

F. Admin Panel and Analytics

The Admin Panel provides advanced control over platform monitoring. Built with reusable components and connected to Supabase in real time, the admin interface allows:

- File Tracking total users and file uploads
- Viewing conversion statistics
- Managing access roles and flagged activity
- Exporting analytics reports (planned feature)

This admin capability ensures the platform remains transparent and manageable during usage by a growing user base.

IV. Results and Evaluation

The Neuron platform was evaluated through systematic testing across multiple criteria including functional correctness, responsiveness, cross-platform stability, and simulation accuracy. Since the system is designed as a simulation environment rather than a physical storage processor, the evaluation emphasizes user interaction quality, system performance, and alignment with the theoretical principles of DNA, graphene, and neural-based storage.

A. Functional Testing:

Neuron was validated for both its visual feedback and backend encoding logic. All major modules were tested against expected behaviors in local and production environments. Functional tests confirmed that:

- Users could upload files of various types (PDF, TXT, MP4, etc.).
- For each conversion type, the system successfully transformed uploaded files into structured symbolic representations.
 - DNA Simulation: Generated valid synthetic nucleotide sequences derived from the binary file content using a 2-bit encoding map.
 - Graphene Simulation: Produced resistance-based binary encodings simulating atomic-scale memory cell behavior.
 - Neural Simulation: Converted input into base64- encoded, timestamped synthetic EEG signals with confidence ratings and waveform structure.
- The progress modal, status icons, and Lottie animations displayed the simulation stages as intended.
- Files could be shared between authenticated users, and Admins had access to user and order analytics via the Admin Panel.
- Table I outlines the main functionalities and their observed results during testing.

Table 1

Summary of Functional Testing Result

Feature	Expected Behavior	Result
File Upload	Accept multiple file types	Success
DNA Encoding Simulation	Animate helical DNA visualization	Success
Graphene Encoding	Display 2D atomic matrix	Success

Neural Encoding	Simulate brainwave patterns	Success
Role-Based Redirection	Redirect unauthorized access	Success
Admin Analytics Panel	Show user stats and file logs	Success
User File Sharing	Enable secure file access between users	Success

B. Cross-Platform Compatibility:

The Neuron was packaged and tested using Tauri.js on the following operating systems:

- Windows 10 / 11
- macOS Ventura
- Ubuntu 22.04 LTS

All platforms successfully supported core features with minor variation in animation rendering speed. File system permissions were properly sandboxed using Tauri's secure APIs, and Supabase ensured consistent backend behavior across environments. No critical bugs were encountered during build or runtime deployment.

C. Performance Metrics:

The system was evaluated for responsiveness, animation smoothness, and load times under moderate usage conditions. The following metrics were recorded:

Initial Load Time: ~ 2.1 seconds (Next.js optimized)

- *Conversion Animation FPS:* 45–60 fps on mid-range hardware
- *Memory Usage (Tauri App):* ~ 85–120 MB at runtime
- *Upload to Simulation Latency:* <2 seconds in 90% of cases

These results demonstrate that Neuron runs efficiently on consumer-grade hardware, making it suitable for demonstrations, classroom use, or personal experimentation.

V. Challenges and Limitations

While Neuron successfully simulates futuristic storage paradigms and delivers an interactive educational experience, several challenges and limitations were encountered throughout development. These limitations highlight areas for future optimization and extension of the platform.

A. Rendering Performance and Animation Complexity

One of the most significant challenges involved managing complex animations for real-time simulation, particularly on lower-end hardware. Smooth rendering of DNA helices, atomic graphene grids, and neural waveform signals required balancing visual quality with performance efficiency. Despite using optimized animation libraries such as Lottie and Framer Motion, frame rate fluctuations were observed on devices with limited GPU acceleration.

B. Platform Compatibility and Build Constraints

Cross-platform support was achieved using Tauri.js, but platform-specific issues arose, particularly in file path resolution and permission handling between Windows, macOS, and Linux systems. File sandboxing and local storage access had to be carefully managed to maintain security without compromising functionality. Future work may involve containerizing parts of the app to enhance portability.

C. Simulation Accuracy and Algorithmic Representation

Neuron integrates simplified encoding algorithms for DNA, graphene, and brain signal formats, transforming binary input into symbolic representations using biologically and technically inspired models. These include 2-bit DNA nucleotide mapping, resistance-based graphene encoding, and synthetic EEG waveform generation.

While these methods enhance educational and conceptual accuracy, they remain purely software-driven. No physical synthesis or interaction with biological or quantum materials occurs.

D. Absence of Offline Mode

The current implementation depends heavily on cloud-based services like Supabase for authentication and storage. This limits functionality in offline or bandwidth-constrained environments. Although desktop packaging reduces some dependence on network availability, full offline simulation support remains an area for future development.

VI. Conclusion and Future Work

In this paper, we presented the development and evaluation of Neuron, a novel simulation platform that encodes binary data into three futuristic formats: DNA sequences, graphene resistance grids, and synthetic brain signals. We detailed the system architecture, conversion methodology, encoding algorithms, and user interaction design, and explained how each module draws inspiration from emerging research in biotechnology, nanotechnology, and cognitive computing. The platform goes beyond visual simulation by implementing real encoding logic that produces structured outputs, potentially usable in future physical data storage systems.

Through rigorous modular design, responsive UI implementation, and cross-platform deployment using Tauri.js and Supabase, Neuron enables users to visualize, understand, and experiment with advanced data storage paradigms in a safe, accessible environment. Functional testing, real-time analytics, and role-based access control ensure a reliable experience across multiple user types.

While the platform is currently simulation-only and lacks integration with physical synthesizers or neural interfaces, its extensibility and encoding framework position it as a future-ready prototyping tool. Planned improvements include offline simulation support, hardware interfacing, format export for synthesis labs, and full bidirectional encoding-decoding flows.

This research contributes a working prototype and framework for educational and experimental interaction with next-generation storage technologies—bridging theory and simulation in a tangible, user-centric way.

References:

References

1. Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. IDC iView, 1142(2011), 1-12.
2. Agneeswaran, V. S., Tonpay, P., & Tiwary, J. (2013). Paradigms for realizing machine learning algorithms. *Big Data*, 1(4), 207-214.
3. Khan, M. A., Khan, S. U. R., Rehman, H. U., Aladhadh, S., & Lin, D. (2025). Robust InceptionV3 with Novel EYENET Weights for Di-EYENET Ocular Surface Imaging Dataset: Integrating Chain Foraging and Cyclone Aging Techniques. *International Journal of Computational Intelligence Systems*, 18(1), 204.
4. Yang, H., Khan, S. U. R., Bilal, O., Chen, C., & Zhao, M. (2025). CEOE-Net: Chaotic Evolution Algorithm-Based Optimized Ensemble Framework Enhanced with Dual-Attention for Alzheimer's Diagnosis. *Computer Modeling in Engineering & Sciences*, 145(2), 2401.
5. Meeran, M. T., Raza, A., & Din, M. (2018). Advancement in GSM Network to Access Cloud Services. *Pakistan Journal of Engineering, Technology & Science [ISSN: 2224-2333]*, 7(1).
6. Maqsood, H., & Khan, S. U. R. (2025). MeD-3D: A multimodal deep learning framework for precise recurrence prediction in clear cell renal cell carcinoma (ccRCC). *Expert Systems with Applications*, 130174.
7. Khan, S. U. R., Asim, M. N., Vollmer, S., & Dengel, A. (2025). Dynamic Weight Adjustment for Knowledge Distillation: Leveraging Vision Transformer for High-Accuracy Lung Cancer Detection and Real-Time Deployment. arXiv preprint arXiv:2510.20438.
8. Singh, D., & Reddy, C. K. (2015). A survey on platforms for big data analytics. *Journal of big data*, 2(1), 1-20.
9. Khan, S. U. R., Zhao, M., & Li, Y. (2025). Detection of MRI brain tumor using residual skip block based modified MobileNet model. *Cluster Computing*, 28(4), 248.
10. Khan, S. U. R., Asim, M. N., Vollmer, S., & Dengel, A. (2025). FloraSyntropy-Net: Scalable Deep Learning with Novel FloraSyntropy Archive for Large-Scale Plant Disease Diagnosis. arXiv preprint arXiv:2508.17653.
11. Raza, A., & Meeran, M. T. (2019). Routine of Encryption in Cognitive Radio Network. *Mehran University Research Journal of Engineering and Technology [p-ISSN: 0254-7821, e-ISSN: 2413-7219]*, 38(3), 609-618.
12. Bilal, O., Hekmat, A., & Khan, S. U. R. (2025). Automated cervical cancer cell diagnosis via grid search-optimized multi-CNN ensemble networks. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 14(1), 67.
13. Emani, C. K., Cullot, N., & Nicolle, C. (2015). Understandable big data: a survey. *Computer science review*, 17, 70-81.
14. Khan, S. U. R. (2025). Multi-level feature fusion network for kidney disease detection. *Computers in Biology and Medicine*, 191, 110214.
15. Mahmood, F., Abbas, K., Raza, A., Khan, M. A., & Khan, P. W. (2019). Three Dimensional Agricultural Land Modeling using Unmanned Aerial System (UAS). *International Journal of Advanced Computer Science and Applications (IJACSA) [p-ISSN : 2158-107X, e-ISSN : 2156-5570]*, 10(1).
16. Khan, S. U. R., Asif, S., Bilal, O., & Rehman, H. U. (2025). Lead-cnn: lightweight enhanced dimension reduction convolutional neural network for brain tumor classification. *International Journal of Machine Learning and Cybernetics*, 1-20.
17. Q. Zhanga, L. T. Yang, Z. Chenc and P. Li, "A survey on deep learning for big data," *Information Fusion*, vol. 42, pp. 146-157, 2018.
18. HUSSAIN, S., Raza, A., MEERAN, M. T., IJAZ, H. M., & JAMALI, S. (2020). Domain Ontology Based Similarity and Analysis in Higher Education. *IEEEP New Horizons Journal*, 102(1), 11-16.

19. Khan, S. U. R., Asif, S., & Bilal, O. (2025). Ensemble Architecture of Vision Transformer and CNNs for Breast Cancer Tumor Detection From Mammograms. *International Journal of Imaging Systems and Technology*, 35(3), e70090.
20. Khan, S. U. R., & Khan, Z. (2025). Detection of Abnormal Cardiac Rhythms Using Feature Fusion Technique with Heart Sound Spectrograms. *Journal of Bionic Engineering*, 1-20.
21. Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research. *Information systems research*, 25(3), 443-448.
22. S. ur R. Khan, Asif. Raza, Muhammad Tanveer Meeran, and U. Bilhaj, "Enhancing Breast Cancer Detection through Thermal Imaging and Customized 2D CNN Classifiers", *VFAST trans. softw. eng.*, vol. 11, no. 4, pp. 80-92, Dec. 2023.
23. Dykes, B. (2017). Big data: Forget volume and variety, focus on velocity. *Forbes*, <https://www.forbes.com/sites/brentdykes/2017/06/28/big-data-forget-volume-andvariety-focus-on-velocity>.
24. O. Bilal, Asif Raza, S. ur R. Khan, and Ghazanfar Ali, "A Contemporary Secure Microservices Discovery Architecture with Service Tags for Smart City Infrastructures ", *VFAST trans. softw. eng.*, vol. 12, no. 1, pp. 79-92, Mar. 2024
25. Rialti, R., Marzi, G., Ciappei, C., & Busso, D. (2019). Big data and dynamic capabilities: a bibliometric analysis and systematic literature review. *Management Decision*.
26. Khan, S. U. R., Asif, S., Zhao, M., Zou, W., Li, Y., & Li, X. (2025). Optimized deep learning model for comprehensive medical image analysis across multiple modalities. *Neurocomputing*, 619, 129182.
27. Braganza, A., Brooks, L., Nepelski, D., Ali, M., & Moro, R. (2017). Resource management in big data initiatives: Processes and dynamic capabilities. *Journal of Business Research*, 70, 328-337.
28. Khan, S. R., Asif Raza, Inzamam Shahzad, & Hafiz Muhammad Ijaz. (2024). Deep transfer CNNs models performance evaluation using unbalanced histopathological breast cancer dataset. *Lahore Garrison University Research Journal of Computer Science and Information Technology*, 8(1).
29. Rahmati, V. (2016). Big data: Now and then. *International Journal of Emerging Computing Methods in Engineering (IJECEME)*, 1(2).
30. M. Wajid, M. K. Abid, A. Asif Raza, M. Haroon, and A. Q. Mudasar, "Flood Prediction System Using IOT & Artificial Neural Network", *VFAST trans. softw. eng.*, vol. 12, no. 1, pp. 210-224, Mar. 2024.
31. Raza, Asif , Soomro, M. H., Shahzad, I., & Batool, S. (2024). Abstractive Text Summarization for Urdu Language. *Journal of Computing & Biomedical Informatics*, 7(02).
32. Al-Khasawneh, Mahmoud Ahmad, Asif Raza, Saif Ur Rehman Khan, and Zia Khan. "Stock Market Trend Prediction Using Deep Learning Approach." *Computational Economics* (2024): 1-32
33. Maltby, D. (2011, October). Big data analytics. In *74th Annual Meeting of the Association for Information Science and Technology (ASIST)* (pp. 1-6).
34. Asif Raza, Salahuddin, Ghazanfar Ali, Muhammad Hanif Soomro, Saima Batool, "Analyzing the Impact of Artificial Intelligence on Shaping Consumer Demand in E-Commerce: A Critical Review", *International Journal of Information Engineering and Electronic Business(IJIEEB)*, Vol.17, No.5, pp. 42-61, 2025.
35. S. U. R. Khan, A. Raza, I. Shahzad and G. Ali, "Enhancing Concrete and Pavement Crack Prediction through Hierarchical Feature Integration with VGG16 and Triple Classifier Ensemble," *2024 Horizons of Information Technology and Engineering (HITE)*, Lahore, Pakistan, 2024, pp. 1-6.
36. Almeida, F. (2018). Big data: concept, potentialities and vulnerabilities. *Emerging Science Journal*, 2(1), 1-10.
37. Bilal, Omair, Arash Hekmat, Inzamam Shahzad, Asif Raza, and Saif Ur Rehman Khan. "Boosting Machine Learning Accuracy for Cardiac Disease Prediction: The Role of Advanced Feature Engineering and Model Optimization." *The Review of Socionetwork Strategies* (2025): 1-30.

38. Khan, Zia, Saif Ur Rehman Khan, Omair Bilal, Asif Raza, and Ghazanfar Ali. "Optimizing Cervical Lesion Detection Using Deep Learning with Particle Swarm Optimization." In 2025 6th International Conference on Advancements in Computational Sciences (ICACS), pp. 1-7. IEEE, 2025.
39. Asif Raza, Inzamam Shahzad, Ghazanfar Ali, and Muhammad Hanif Soomro. "Use Transfer Learning VGG16, Inception, and Reset50 to Classify IoT Challenge in Security Domain via Dataset Bench Mark." *Journal of Innovative Computing and Emerging Technologies* 5, no. 1 (2025).
40. Khan, Saif Ur Rehman, Asif Raza, Inzamam Shahzad, and Shehzad Khan. "Subcellular Structures Classification in Fluorescence Microscopic Images." *International Conference on Computing & Emerging Technologies*, pp. 271-286. Cham: Springer Nature Switzerland, 2023.
41. Al Nuaimi, A. J. J., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities *Journal of Internet Services and Applications*.
42. Hurwitz, J., Nugent, A., Halper, F., & Kaufman, M. (2013). *Big data for dummies* (Vol. 336). Hoboken, NJ: John Wiley & Sons.
43. Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE access*, 2, 652-687.
44. Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future, 2007(2012)*, 1-16.
45. Bellaachia A, Guven E (2003) Predicting breast cancer survivability using data mining techniques. *J Soc Ind Appl Math* 7(1):37-42
46. Chi, C. L., Street, W. N., & Wolberg, W. H. (2007). Application of artificial neural network-based survival analysis on two breast cancer datasets. In *AMIA annual symposium proceedings* (Vol. 2007, p. 130). American Medical Informatics Association.
47. Raza, Asif, Inzamam Shahzad, Muhammad Salahuddin, and Sadia Latif. "Satellite Imagery Employed to Analyze the Extent of Urban Land Transformation in The Punjab District of Pakistan." *Journal of Palestine Ahliya University for Research and Studies* 4, no. 2 (2025): 17-36.
48. Buciński, A., Bączek, T., Krysiński, J., Szoszkiewicz, R., & Załuski, J. (2007). Clinical data analysis using artificial neural networks (ANN) and principal component analysis (PCA) of patients with breast cancer after mastectomy. *Reports of Practical Oncology and Radiotherapy*, 12(1), 9-17.
49. Joshi, J., Doshi, R., & Patel, J. (2014). Diagnosis of breast cancer using clustering data mining approach. *International Journal of Computer Applications*, 101(10), 13-17.
50. Saleema, J. S., Bhagawathi, N., Monica, S., Shenoy, P. D., Venugopal, K. R., & Patnaik, L. M. (2014). Cancer prognosis prediction using balanced stratified sampling. *arXiv preprint arXiv:1403.2950*.
51. Khosravian, A., & Ayat, S. (2016). Diagnosing breast cancer type by using probabilistic neural network in decision support system. *Int J Knowl Eng*, 2(1), 73-6.
52. Waqas M, Khan Z, Anjum S, Tahir MA. Lung-Wise Tuberculosis Analysis and Automatic CT Report Generation with Hybrid Feature and Ensemble Learning. In *CLEF (Working notes) 2020 Sep* (pp. 1-10).
53. Salama, G. I., Abdelhalim, M., & Zeid, M. A. E. (2012). Breast cancer diagnosis on three different datasets using multi-classifiers. *Breast Cancer (WDBC)*, 32(569), 2.
54. Medjahed, S. A., Saadi, T. A., & Benyettou, A. (2013). Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules. *International Journal of Computer Applications*, 62(1).
55. Kandhasamy, J. P., & Balamurali, S. J. P. C. S. (2015). Performance analysis of classifier models to predict diabetes mellitus. *Procedia Computer Science*, 47, 45-51.
56. Manjisha, M., & Hari Kumar, R. (2016). Performance analysis of KNN and K-means clustering for robust classification of epilepsy from EEG signals. In *Int Conf Wirel Common Signal Process Netw (NISPNET)*.

57. Benedetti, F., Beneventano, D., Bergamaschi, S., & Simonini, G. (2019). Computing inter-document similarity with context semantic analysis. *Information Systems*, 80, 136-147.
58. Abawajy, J. (2015). Comprehensive analysis of big data variety landscape. *International journal of parallel, emergent and distributed systems*, 30(1), 5-14.
59. Waqas M, Tahir MA, Qureshi R. Ensemble-based instance relevance estimation in multiple-instance learning. In 2021 9th European workshop on visual information processing (EUVIP) 2021 Jun 23 (pp. 1-6). IEEE.
60. Habibpour, R., & Khalilpour, K. (2014). A new hybrid k-means and k-nearest-neighbor algorithms for text document clustering. *International Journal of Academic Research*, 6(3).
61. Kuśmirek, W., Szmurło, A., Wiewiórka, M., Nowak, R., & Gambin, T. (2019). Comparison of kNN and k-means optimization methods of reference set selection for improved CNV callers' performance. *BMC bioinformatics*, 20(1), 1-10.
62. Mittal, K., Aggarwal, G., & Mahajan, P. (2019). Performance study of K-nearest neighbor classifier and K-means clustering for predicting the diagnostic accuracy. *International Journal of Information Technology*, 11(3), 535-540.
63. Waqas M, Tahir MA, Al-Maadeed S, Bouridane A, Wu J. Simultaneous instance pooling and bag representation selection approach for multiple-instance learning (MIL) using vision transformer. *Neural Computing and Applications*. 2024 Apr;36(12):6659-80.
64. Zhao, D., Hu, X., Xiong, S., Tian, J., Xiang, J., Zhou, J., & Li, H. (2021). K-means clustering and kNN classification based on negative databases. *Applied Soft Computing*, 110, 107732.
65. Kaiser, M. S., Mahmud, M., Noor, M. B. T., Zenia, N. Z., Al Mamun, S., Mahmud, K. A., & Hussain, A. (2021). iWorkSafe: towards healthy workplaces during COVID-19 with an intelligent pHealth App for industrial settings. *Ieee Access*, 9, 13814-13828.
66. Narayana, M., & Bhavani, N. P. G. (2022, February). Traffic sign identification on rainy conditions using K means algorithm comparison with KNN and SVM. In 2022 International Conference on Business Analytics for Technology and Security (ICBATS) (pp. 1-6). IEEE.
67. Waqas M, Tahir MA, Qureshi R. Deep Gaussian mixture model based instance relevance estimation for multiple instance learning applications. *Applied intelligence*. 2023 May;53(9):10310-25.
68. Gómez, F., Flores, Y., & Vadász, M. (2022). Comparative analysis of the K-nearest-neighbour method and K-means cluster analysis for lithological interpretation of well logs of the Shushufindi Oilfield, Ecuador. *Rudarsko-geološko-naftni zbornik*, 37(4), 155-165.
69. Wang, W., Liu, J., Xia, F., King, I., & Tong, H. (2017, April). Shifu: Deep learning-based advisor-advisee relationship mining in scholarly big data. In Proceedings of the 26th international conference on World Wide Web companion (pp. 303-310).
70. Waqas M, Tahir MA, Khan SA. Robust bag classification approach for multi-instance learning via subspace fuzzy clustering. *Expert Systems with Applications*. 2023 Mar 15;214:119113. Nascimento, D. M., Ferreira, M., & Pardal, M. L. (2022). Does Big Data Require Complex Systems? A Performance Comparison between Spark and Unicage Shell Scripts. *arXiv preprint arXiv:2212.13647*.
71. Waqas M, Ahmed SU, Tahir MA, Wu J, Qureshi R. Exploring multiple instance learning (MIL): A brief survey. *Expert Systems with Applications*. 2024 Sep 15;250:123893.
72. Lamsal, R. (2021). Design and analysis of a large-scale COVID-19 tweets dataset. *Applied Intelligence*, 51, 2790-2804. <https://doi.org/10.1007/s10489-020-02029-z>
73. Lopez, C. E., Gallemore, C., "An Augmented Multilingual Twitter dataset for studying the COVID-19 infodemic" *Soc. Netw. Anal. Min.* 11, 102 (2021). DOI: s13278-021-00825-0

-
74. Lloyd, S. P. Least squares quantization in pcm. *Information Theory, IEEE Transactions on* 28, 2 (1982), 129-137.
75. Waqas M, Bandyopadhyay R, Showkatian E, Muneer A, Zafar A, Alvarez FR, Marin MC, Li W, Jaffray D, Haymaker C, Heymach J. The Next Layer: Augmenting Foundation Models with Structure-Preserving and Attention-Guided Learning for Local Patches to Global Context Awareness in Computational Pathology. arXiv preprint arXiv:2508.19914. 2025 Aug 27.

