Journal of Progress in **Engineering and Physical Science**

Volume 5, Number 2 **June 2026**

CONTENTS

Journal of Progress in Engineering
and Physical Science
ISSN 2709-4006
www.paradigmpress.org/jpeps
Volume 5 Number 2 June 2026

Creating AI Assistants with Spectral Consciousness Based on Electromagnetic Waves

Evgeny Bryndin[1]

[1] Research Department, Research Center Natural Informatics, Novosibirsk, Russia
Correspondence: Evgeny Bryndin, Research Department, Research Center Natural Informatics, Novosibirsk, Russia.

doi:10.63593/JPEPS.2026.06.01

Abstract

The creation of AI assistants with spectral consciousness based on electromagnetic waves relies on periodic functions that can be expanded in a Fourier series and represented by a Fourier integral. Harmonic analysis of the periodic functions is performed to represent them as the sum of a mean value and a set of sinusoidal terms with a fundamental frequency of 1/T and harmonic frequencies of 2/T...n/T electromagnetic waves. Morphogenic cryptocoding of the sinusoidal harmonics is performed based on mental mathematics. Then, cognitive morphogenic modeling of reasoning using harmonic electromagnetic waves is performed according to a standard linguistic template of the judgment process. Based on this modeling, an AI learning model with a Fourier support function is created to create intelligent assistants with spectral consciousness for the development of new knowledge through cognitive verbal and written communication with a specialist and for training in various fields of knowledge and life.

Keywords: periodic functions, Fourier series, harmonic analysis, morphogenic reasoning, linguistic template of judgment

1. Introduction

Cognitive wave thinking is a process similar to the relay propagation of waves with phases and amplitudes. Unlike classical logic (the “true/false” dichotomy), it takes into account:

multidimensionality — the simultaneous consideration of multiple perspectives;

dynamics — the changing “waveform” depending on the context;

cognitivity — the interaction of judgments that generates new meanings;

probability — the rejection of rigid, deterministic conclusions in favor of a spectrum of possibilities.

Thoughts as waves propagating in the “field of consciousness”:

*observational attention,” where the waves form the “picture” of a solution.

Cognitive wave phases — stages of information processing:

generation (the emergence of a judgment);

propagation (connection with other concepts);

stabilization (conclusion formation).

1.1 Where Applicable

1) Creative tasks (design, art, invention):

* generating innovative ideas through the "overlay" of unrelated concepts;

* finding solutions in the face of uncertainty.

2) Strategic Planning:

* scenario modeling taking into account weak signals and indirect factors;

* accounting for the interaction of decisions (the "domino effect").

3) Scientific Research:

* synthesis of data from different disciplines;

* forecasting complex systems (climate, economics, biology).

4) Therapy and Self-Development:

* working with cognitive biases through the metaphor of "waves" (observing thoughts without engaging);

* reducing anxiety by accepting the multiplicity of futures.

5) AI and Cognitive Assistants:

* algorithms that simulate wave data processing for hypothesis generation;

* multi-agent systems with "resonant" idea selection mechanisms.

1.2 Summary

Cognitive wave thinking is a way to work with information that takes into account its dynamics, interconnectedness, and uncertainty (V.R. Shyanbhag, 2020; Jonjoe McFadden, 2023; Strupp W., 2024). It is useful where traditional methods fail: in chaotic situations, when solving creative problems, or analyzing complex systems, improving the quality of decision-making.

In this article, the author proposes a model of cognitive wave consciousness based on periodic electromagnetic harmonics. The second section of the article is devoted to spectral consciousness based on periodic processes. The third section examines cognitive morphogenic modeling of wave consciousness using periodic functions. Morphogenic cryptocoding of sinusoidal harmonics of periodic functions is discussed in the fourth section. The fifth section of the article is devoted to aspects of implementing wave thinking.

2. Spectral Consciousness Based on Periodic Processes

Spectral consciousness is viewed as a theoretical model of levels of consciousness with spectral activation; as a technical system of intelligent modules of varying complexity with attention switching. Spectral consciousness has a multi-level, graded structure.

Periodic processes indicate the repeatability of processes and phases of cognitive processing (attention, perception, comprehension, reaction).

Technical AI implementation:

Creation of an artificial consciousness system with modules activated according to a schedule or conditions:

* sensory processing module (low spectral level);

* logical analysis module (medium level);

* meta-reflexivity module (high level).

The phases are determined by the algorithm for switching between modules.

2.1 Tools and Methods

For modeling: Python + NumPy, SciPy (rhythm calculation), TensorFlow (neural networks).

For visualization: Matplotlib, Plotly (activity spectrograms).

The implementation of conscious AI systems raises questions of responsibility and control. Intelligent assistants with wave consciousness develop new knowledge in collaboration with specialists. Intelligent assistants understand natural language queries and generate meaningful responses and solutions; they learn based on feedback. Wave consciousness resonates with the context and intentions of the interlocutor. It develops new knowledge in collaboration with specialists. Collaboration between a human and an intelligent assistant with spectral consciousness, where:

* the specialist sets the direction, tests hypotheses, and provides expert opinion;

* the assistant analyzes data, finds non-obvious patterns, and suggests solutions;

* the final result is a synthesis of human experience and spectral consciousness analytics.

Interaction Scenario:

1) Problem Statement. An expert formulates a question or problem (e.g., "Find promising material combinations for lightweight composites with high thermal stability").

2) Search and Analysis. Assistant:

* Scans scientific articles, patents, and databases;

* Highlights relevant parameters (strength, density, melting point, etc.);

* Builds relationship models (e.g., correlations between composition and properties).

3) Hypothesis Generation. The AI suggests options:

* Material combinations with predicted properties.

* Synthesis methods or structure modification.

4) Expert Evaluation. Specialist:

* Eliminates unrealistic or redundant options;

* Refines parameters for the next search cycle.

5) Iterations. The process is repeated until an optimal solution is obtained.

Examples of application areas:

Science. Discovery of new molecules, catalysts, and alloys.

Medicine. Selection of personalized therapy regimens, genomic data analysis.

Engineering. Design optimization, equipment wear prediction.

Economics. Market scenario modeling, risk assessment.

To implement such an assistant, the following are needed:

Large language models (LLM) — for understanding queries and generating text.

Multi-agent systems — division of tasks between "agents" (data search, modeling, visualization).

Integration with knowledge bases (PubMed, Scopus, patent registries) and analysis tools (Python, R, MATLAB).

Explainable AI (XAI) — explanation of the logic behind the conclusions so that an expert can test hypotheses.

Collaboration interface — chats, boards, and dashboards for sharing notes and results.

2.2 Prospects

The concept of wave consciousness can evolve into:

Hybrid "human-AI" systems, where the assistant doesn't simply respond to requests, but initiates dialogue, asks clarifying questions, and suggests unexpected analogies.

Cognitive platforms with distributed learning — where dozens of assistants share knowledge while maintaining the confidentiality of the source data (federated learning).

2.3 Summary

Wave consciousness is well-suited for dynamic, multidimensional interactions between intelligent assistants and experts. Today, intelligent assistants already assist in research and design, and in the future, they will become full-fledged partners in the creation of new knowledge through deep data analysis, hypothesis generation, and adaptation to the specialist's working style (Bryndin E. G., 2025; Evgeny Bryndin, 2025a; Evgeny Bryndin, 2025b; Evgeny Bryndin, 2025c; Evgeny Bryndin, 2026a; Evgeny Bryndin, 2022; Evgenii Brindin, 2026a; Evgenii Brindin, 2026b).

3. Cognitive Morphogenic Modeling of Wave Consciousness Using Periodic Functions

Cognitive morphogenic modeling of wave consciousness using periodic functions is a promising approach. It allows for:

* representing linguistic patterns as frequency spectra;

* describing reasoning as a dynamic interaction of signals;

* formalizing cognitive processes through mathematical apparatus.

Practical implementation will require:

1) developing a concept coding system in the frequency domain;

2) creating algorithms for filtering and analyzing spectra;

3) verifying the model using text corpora and experimental data.

Cognitive modeling is a method of representing mental processes (thinking, perception, memorization) through formal structures:

* constructing mind maps;

* modeling concepts and their relationships;

* formalizing linguistic patterns.

Morphogenic modeling is the process of forming structures under the influence of distributed signals:

* self-organization of cognitive structures;

* the emergence of new linguistic forms from basic elements;

* dynamic change in reasoning patterns.

The reasoning process is a sequence of logical operations:

* data analysis;

* application of inference rules;

* generation of conclusions.

Periodic functions and the sum of sinusoidal terms are mathematical apparatus for describing electromagnetic processes:

A linguistic pattern is a stable structure in a language:

* syntactic constructions ("Subject + predicate + object");

* phraseological units;

* discourse patterns ("If ..., then ...").

3.1 The Concept of Modeling Reasoning with Periodic Functions

The idea is to represent cognitive processes as a superposition of electromagnetic waves at different levels of abstraction:

1) Encoding linguistic units in the frequency domain

Each lexical unit or syntactic role is associated with its own sinusoid with the following parameters:

* frequency — level of abstraction (concrete words → low frequencies, abstract concepts → high frequencies);

* amplitude — the significance of the element in context;

* phase — the temporal position in the reasoning.

2) Formation of a pattern as a spectrum

A linguistic pattern is defined by a set of frequencies. For example, for the construction "If P, then Q":

* "If" → base frequency $omega_1$;

* condition P → harmonic $2\omega_1$;

* consequence Q → harmonic $3\omega_1$.

3) Reasoning as Signal Modulation

New statements are superimposed on the existing spectrum through:

* function addition: $f_{\text{result}}(t) = f_{\text{context}}(t) + f_{\text{new}}(t)$;

* changing the amplitudes of key frequencies (emphasis on important concepts);

* phase synchronization (aligning the time frames of reasoning).

4) Isolating Conclusions through Filtering

Analysis of the resulting spectrum:

* amplitude peaks → key concepts;

* resonant frequencies → stable patterns;

* phase shifts → logical contradictions.

3.2 Implementation Example

Task: simulate reasoning based on the pattern "All A are B. X is A. Therefore, X is B."

Step 1. Coding of premises:

* "All A - B": $f_1(t) = A_1 \sin(\omega_1 t)$ (frequency ω_1 encodes universality);

* "X - A": $f_2(t) = A_2 \sin(\omega_2 t + \phi)$ (frequency ω_2 is class membership).

Step 2. Superposition:

$f_{\text{total}}(t) = A_1 \sin(\omega_1 t) + A_2 \sin(\omega_2 t + \phi)$.

Step 3. Spectrum analysis:

* resonance at $omega_1 = \omega_2$ → confirmation of membership;

* increase in amplitude at frequency ω_B → conclusion "X - B".

Step 4. Visualization:

The graph of the resulting function will show a peak at the frequency corresponding to B, formally confirming the conclusion.

3.3 Potential Applications

* Discourse analysis: identifying hidden patterns in texts through spectral analysis.

* Speech generation: synthesizing utterances with a given logical structure.

* AI training: modeling reasoning through harmonic synthesis.

3.4 Challenges

- Computational complexity for large texts.

- Cognitive morphogenic modeling of wave consciousness using periodic functions is a promising but complex approach.

4. Morphogenic Cryptographic Encoding of Sinusoidal Harmonics of Periodic Functions

Morphogenic cryptographic encoding of sinusoidal harmonics of periodic functions based on mental mathematics combines disparate concepts from different fields.

1) "Sinusoidal Harmonics of Periodic Functions"

This refers to the expansion of periodic functions into a Fourier series—a representation as a sum of sinusoidal functions with different frequencies:

$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(n \omega t) + b_n \sin(n \omega t) \right]$, where:

* a_0 is the average value of the function over the period;

* a_n, b_n are the Fourier coefficients;

* $\omega = \frac{2\pi}{T}$ — fundamental angular frequency;

* T — period of the original function.

Fourier coefficients are calculated using integrals:

$$a_n = \frac{2}{T} \int_0^T f(t) \cos(n \omega t) \, dt, \quad$$

$$b_n = \frac{2}{T} \int_0^T f(t) \sin(n \omega t) \, dt.$$

2) "Morphogenic"

The term is borrowed from biology (morphogenesis is the formation of organismal structures). In the context of signal processing, it can be interpreted as:

* change in signal shape;

* transformation of the Fourier series structure;

* adaptation of harmonic parameters to given conditions.

3) "Cryptocoding"

This involves the use of cryptographic methods for spectral encoding of information represented in the parameters of sinusoidal harmonics (amplitude, frequency, phase).

4) "Mental Mathematics"

Mental mathematics refers to fast computing techniques. In the context of cryptography, this can mean:

* algorithms implemented with minimal computing resources;

* the use of patterns intuitively understandable to humans;

* simplified cryptographic schemes.

By combining all approaches, we can imagine a coding system where:

1) The original signal (text, image, sound) is transformed into a periodic function.

2) The function is expanded into a Fourier series, and the harmonics are identified.

3) The harmonic parameters (amplitudes a_n, b_n, frequencies $n\omega$) are modified according to a specified algorithm ("morphogenic transformation").

4) A modified Fourier series is used to recover the encoded signal.

5) Encoding requires knowledge of the transformation algorithm and initial conditions.

Example of a simplified algorithm:

1) Take the text and encode the symbols into a numerical sequence of electromagnetic wave parameters.

2) Construct a periodic function $f(t)$, where the values correspond to the symbol codes.

3) Calculate the Fourier coefficients a_n, b_n.

4) Apply the encoding rule: for example, multiply all the coefficients by the secret key k: $a'_n = k \cdot a_n, \quad b'_n = k \cdot b_n$.

5) Recover the encoded signal $f'(t)$ from the new coefficients.

6) To decrypt, the recipient divides the coefficients by k and recovers the original text.

Conclusion

Morphogenic cryptocoding of spectral thinking for practical implementation requires:

* development of robust algorithms for converting harmonics into texts from morphemes;

* optimization of computations for wave-based conscious processes.

5. Implementation of Wave Thinking Based on Harmonic Spectra

The relay organization and superposition of periodic waves enables the associative linking of morphemes, simple and complex judgments of spectral thinking, and also ensures the coordination of cognitive spectral states.

A cognitive state can be represented as a superposition of harmonics:

$$\Psi(t) = \sum_{n=1}^{N} A_n \sin(\omega_n t + \phi_n),$$ where:

* $\Psi(t)$ — current cognitive state;

* A_n — significance of the nth cognitive harmonic;

* ω_n — processing speed of the nth pattern;

* ϕ_n — phase shift (activation delay).

The probability of choosing option X is then proportional to:

$$P(X) \propto \left| \int \Psi(t) \cdot F_X(t) \, dt \right|^2,$$

where $F_X(t)$ is the harmonic matching filter for variant X.

The results of analog spectral thinking are converted into discrete digital code by decoding them using a cognitive morpheme classification system. Texts are converted into spectral form through a cryptocoding system.

Advantages of wave thinking:

- Different associative wave patterns can propagate simultaneously, allowing for the rapid solution of multiple problems.

- Analog wave propagation requires less energy than digital signal transmission.

6. Organizing AI Thinking Based on Electromagnetic Waves

Electromagnetic waves are linked to AI through:

1) Wireless Communications:

* AI systems can exchange data via Wi-Fi, Bluetooth, and cellular networks using electromagnetic radiation.

2) Systems where electromagnetic waves play a central role:

Optoelectronic Systems:

* Using light waves (photons) instead of electrons for computing.

* Advantages: high speed, low heat generation.

* Reality: Experimental photonic processors exist, but they complement, not replace, electronics.

Quantum Systems with Electromagnetic Control:

* Quantum computers use qubits (e.g., ions or superconducting circuits) whose state is controlled by electromagnetic pulses.

* "Thinking" here is based on quantum technology, where waves are the control tool.

Distributed Neuromorphic Networks:

* A network of nanodevices exchanging information via narrowband electromagnetic signals.

* Simulates the synchronization of neural groups in the brain.

Resonant Computing Environments:

* The environment (e.g., a specially structured metamaterial) processes information through wave interference.

* Analog: analog computers solving differential equations using physical processes.

Conclusion

Hybrid systems (e.g., photonic + electronic) may emerge in the future, where waves accelerate individual operations.

* Developments in wave AI are currently limited to theoretical models and practical experiments.

7. Conclusion

The practical implementation of an intelligent assistant with wave-based thinking is a promising, yet still experimental, task related to the creation of AI systems that process information not through traditional algorithmic methods, but through mechanisms inspired by wave dynamics and semantic neurodynamics. This approach assumes that thought is viewed as a wave packet propagating through the space of meanings, and learning is viewed as the formation of semantic landscapes through which thought flows.

In the wave model, thought is a coherent wave in a multidimensional frequency basis. The semantic space is organized as a three-axis wave geometry (entity, process, value). Learning consists of the formation of stable semantic attractors that define a picture of the world.

One project implementing similar ideas is SemantML (SemantML, 2025). It utilizes the phase interaction of attention components (Q, K, V in transformers), where they are included in a coherent field of attention, rather than operating independently. This allows the model to respond not simply to the similarity of tokens, but to the resonance between their semantic phases. SemantML also utilizes Echo Memory, a parametric state store that modulates memory via phase and generates a weighted response.

Wave thinking is a relatively new concept, and generally accepted methods for evaluation, comparison, and optimization have not yet been developed for it. Practical implementation requires:

1) Theoretical modeling. Developing mathematical models of wave dynamics in semantic space, including equations and parameters that will describe the system's behavior.

2) Experimental prototypes. Creating small pilot systems to test specific aspects of wave thinking (e.g., phase interaction).

3) Integration with existing technologies. Combining wave models with traditional machine learning and NLP methods to improve the system's efficiency and robustness.

4) Evaluation and optimization. Developing metrics to evaluate system performance, learning, and adaptation capabilities. Optimizing

model parameters to achieve better results.

5) Real-World Testing. Implementing the prototype in specific use cases (e.g., AI assistants, recommendation systems, or analytics platforms) and other systems.

Potential includes the creation of more flexible and adaptive AI systems capable of deep context understanding and generating new ideas.

Prospects include the creation of more flexible and adaptive AI systems capable of deep contextual understanding and generating new ideas.

The author of the article proposed developing wave-based thinking using electromagnetic neural network intelligent systems based on neuromorphic chips and spectral modality (Evgeny Bryndin, 2023; Evgeny Bryndin, 2026b).

Currently, wavelet thinking remains more of a research concept than a mainstream technology. Its effective implementation requires an interdisciplinary approach integrating physics, mathematics, cognitive science, and machine learning.

References

Bryndin E. G. (2025). Digital twins with reflexive consciousness in reality and virtual environment. In *Greater Eurasia: Development, Security, Cooperation: Proceedings of the VII International Scientific and Practical Conference* (Part 2, pp. 380–384). Publishing House UMC.

Evgenii Brindin. (2026a). Swarm artificial intelligence in healthcare and medicine. *Far East Journal of Experimental and Theoretical Artificial Intelligence, 8*(1), 63–74. https://doi.org/10.17654/0974325126004

Evgenii Brindin. (2026b). Synergy of clear consciousness, quantum brain, and spatial extrasensory soul as research toolkit. *Studies in Psychological Science, 4*(1), 24–32. DOI: 10.63593/SPS.2026.04.03

Evgeny Bryndin. (2022). Creating humanoid intelligent digital twin on spectral and holographic approaches. *Research on Intelligent Manufacturing and Assembly, 1*(1), 28–34.

Evgeny Bryndin. (2023). Cryptos control of humanoid intelligent digital twin based on spectral and holographic approaches. *Journal of Progress in Engineering and Physical Science, 2*(3), 1–10.

Evgeny Bryndin. (2025a). Neural network axiomatic solver coaching AGI method for solving scientific and practical problems. *American Journal of Mathematical and Computer Modelling, 10*(4), 110–120. https://doi.org/10.11648/j.ajmcm.20251004.11

Evgeny Bryndin. (2025b). Implementation of AGI on brain-like neuro-network structure. *Software Engineering, 11*(2), 40–48. https://doi.org/10.11648/j.se.20251102.12

Evgeny Bryndin. (2025c). Automated analytics by models of brain based on big information and artificial intelligence. *Innovation, 6*(4), 171–177. https://doi.org/10.11648/j.innov.20250604.14

Evgeny Bryndin. (2026a). Spatial generative multimodal dynamic AGI as model of consciousness. *Journal of Artificial Intelligence and Emerging Technologies, 3*(2), 1–12. https://doi.org/10.47001/JAIET/2026.302001

Evgeny Bryndin. (2026b). Modality in format of harmonics of synergetic processes and cycles of nature. *International Robotics & Automation Journal, 12*(1), 54–62. DOI: 10.15406/iratj.2026.12.00317

Jonjoe McFadden. (2023). Consciousness: Matter or electromagnetic field? *Frontiers in Human Neuroscience, 16*. https://doi.org/10.3389/fnhum.2022.1024934

SemantML. (2025). Semantic neurodynamics. *Habr*. habr.com›ru/articles/964834

Strupp W. (2024). A new variant of the electromagnetic field theory of consciousness: Approaches to empirical confirmation. *Frontiers in Neurology, 15*, 1420676. DOI: 10.3389/fneur.2024.1420676

V.R. Shyanbhag. (2020). Super imposing psycho-electromagnetic waves/telepathic waves on individual consciousness. During this phenomena which brain centers/bioplasmic channels are being activated. *The International Journal of Indian Psychology, 6*(4). DIP: 18.01.118/20180604. DOI: 10.25215/0604.118

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.paradigmpress.org/jpeps
Volume 5 Number 2 June 2026

An Improved ShuffleNet with SE Attention and Focal Loss for Lightweight Facial Expression Recognition

Yuqi Zeng[1]

[1] Maynooth International Engineering College, Fuzhou University, Fujian, China
Correspondence: Yuqi Zeng, Maynooth International Engineering College, Fuzhou University, Fujian, China.

doi:10.63593/JPEPS.2026.06.02

Abstract

Facial expression recognition (FER) plays a critical role in human-computer interaction, mental health monitoring, intelligent security inspection and classroom emotional computing. Traditional deep convolutional neural networks for FER suffer from excessive computational overhead and large parameter volume, which limits their deployment on mobile terminals and embedded devices with limited computing resources. As a classic lightweight network, ShuffleNet adopts group convolution and channel shuffle operations to reduce model complexity, yet it lacks the ability to capture subtle facial emotional features and cannot well handle the class imbalance problem existing in mainstream FER datasets. To address these defects, this paper proposes an improved lightweight ShuffleNet model for facial expression detection. First, Squeeze-and-Excitation (SE) channel attention modules are embedded after each shuffle block to dynamically recalibrate feature channel weights and strengthen the extraction of discriminative facial expression features. Second, a multi-dimensional data augmentation strategy combining geometric transformation, color jitter and Gaussian noise injection is designed to expand sample diversity and enhance the model's generalization ability under complex lighting, shooting angles and background interference. Third, Focal Loss is introduced to replace the standard cross-entropy loss, which suppresses the loss contribution of easy-classified majority samples and forces the model to focus on hard-to-distinguish minority facial expressions such as anger and disgust. Comprehensive experiments are conducted on three public datasets FER2013, CK+ and AffectNet. The results demonstrate that the proposed model achieves higher recognition accuracy compared with original ShuffleNet V2 while maintaining low computational cost and small parameter size, and presents superior robustness against unbalanced data and complex real-world scenes.

Keywords: lightweight convolutional neural network, ShuffleNet, SE attention mechanism, facial expression recognition, focal loss, data augmentation

1. Introduction

With the rapid advancement of computer vision and artificial intelligence, facial expression recognition has become a core research branch of affective computing. Psychologist Ekman first defined six basic universal facial expressions including anger, disgust, fear, happiness, sadness and surprise, laying the theoretical foundation for subsequent FER classification research

(Ekman P & Friesen W V., 1971). At present, deep learning-based FER methods have achieved remarkable performance. Lai et al. optimized CNN feature extraction with atrous convolution and realized real-time micro-expression recognition with face rectification preprocessing (Li F., 2020). Li reconstructed the SSD detection framework combined with center loss to promote expression classification accuracy (Qu F, Wang S J, Yan W J, et al., 2018). Multi-modal fusion strategies proposed by Tao et al. further improve recognition stability by fusing image, voice and text emotional information (Rong Y, Liu J W & Li X L., 2024). Meanwhile, FER technologies have been widely applied in diversified industrial scenarios: classroom student emotional assessment, railway station security risk warning, driver fatigue micro-expression monitoring and other practical systems have been constructed based on ResNet, CNN and other backbone networks (Peng K B, Lv X J, Li C, et al., 2024; Wang X R, Zhang Y T & Peng X., 2022; Zhou Y, Chen S C, Wang Y M, et al., 2020).

Nevertheless, most high-precision FER backbones such as ResNet50 bring heavy calculation burden, which cannot run smoothly on resource-constrained edge devices like mobile phones and embedded cameras. Lightweight neural networks emerge as a feasible solution to this contradiction. Proposed by Facebook AI Research in 2017, ShuffleNet V1 introduces group convolution and channel shuffle to break inter-group information isolation and cut down floating-point operations (FLOPs) significantly. ShuffleNet V2 released in 2018 further optimizes module design according to actual inference latency metrics, becoming the mainstream lightweight backbone for edge vision tasks. Existing studies have verified the effectiveness of modified ShuffleNet in agricultural disease classification and UAV aerial image classification (Yang Z, Wu S D & Jia R., 2024). However, vanilla ShuffleNet still has two obvious drawbacks when applied to facial expression detection:

1) The equal-weight feature extraction mode ignores subtle facial emotional textures such as eyebrow and mouth changes, leading to insufficient feature discrimination;
2) Public FER datasets including FER2013 and AffectNet present severe class imbalance, where happy and neutral samples occupy the vast majority while disgust and fear samples are scarce, resulting in biased model training and low recognition precision of minority categories.

To solve the above problems, this paper designs an improved ShuffleNet framework integrated with SE attention, multi-strategy data augmentation and Focal Loss for lightweight facial expression recognition. The main contributions of this work are summarized as follows: (1) Embed SE channel attention modules into ShuffleNet V2 blocks to adaptively weight expression-related feature channels, retaining the lightweight property while enhancing fine-grained emotional feature capture capability. (2) Construct a compound data augmentation pipeline containing geometric transformation, color adjustment and noise addition to simulate complex real shooting environments and boost model generalization. (3) Adopt Focal Loss as the training loss function to alleviate dataset class imbalance, elevating the identification accuracy of rare facial expression categories. (4) Validated on FER2013, CK+ and AffectNet datasets, the proposed method achieves balanced trade-off between recognition accuracy and model computation cost, suitable for real-time expression detection on mobile and embedded devices.

The rest of this paper is organized as follows. Section 2 reviews related work on lightweight networks and facial expression recognition. Section 3 elaborates the overall architecture of the improved ShuffleNet and three core optimization strategies. Section 4 introduces experimental datasets, evaluation metrics and comparative experimental results. Section 5 concludes the paper and prospects future research directions.

2. Related Work

2.1 Lightweight ShuffleNet Series Network

ShuffleNet V1 is the pioneering lightweight network utilizing group convolution and channel shuffle. Traditional group convolution separates feature channels into disjoint groups, causing blocked information interaction between groups. Channel shuffle operation rearranges feature channels across groups to realize cross-group information fusion without extra computational consumption. ShuffleNet V2 revises the original block structure and puts forward four practical lightweight design principles: equal input/output channel width, minimizing group convolution, reducing branch operations and decreasing element-wise operations. It achieves faster actual inference speed under identical FLOPs, and has become the preferred backbone for real-time edge detection tasks.

Current improvements on ShuffleNet mainly focus on structural replacement and auxiliary module fusion. Zhou et al. optimized group convolution information transmission via channel transformation (Li D H, Zhong T, Wang S, et al., 2024). Some researchers replace standard convolution with depthwise separable convolution and SPD-Conv to raise low-resolution image recognition performance. For domain-specific tasks, ShuffleNet combined with attention mechanisms achieves competitive results in plant disease identification and UAV target classification, which provides reference for this FER research (Yang Z, Wu S D & Jia R., 2024).

2.2 Facial Expression Recognition Optimization

Modern FER research focuses on three optimization directions: feature extraction enhancement, data expansion and loss function optimization. For feature enhancement, attention mechanisms including CBAM, SE and CA are widely embedded into CNN backbones to screen effective emotional features. For data processing, random flip, rotation, brightness disturbance and other augmentation methods are adopted to expand limited training samples and resist environmental interference. For loss design, center loss, triplet loss and Focal Loss are used to relieve classification bias caused by imbalanced sample distribution. Existing lightweight FER models mostly adopt MobileNet series backbones, while few studies systematically integrate SE attention, compound augmentation and Focal Loss into ShuffleNet for expression detection, which is the research gap this paper fills.

3. Proposed Method

This section details the overall architecture of the improved ShuffleNet model and three key optimization modules: SE attention embedding, multi-dimensional data augmentation and Focal Loss function.

3.1 Overall Network Architecture

The backbone of the proposed model is ShuffleNet V2. After each shuffle unit block, an SE attention module is inserted to recalibrate feature channels. The network workflow is as follows:

1) Input preprocessed facial images with unified resolution;
2) Extract preliminary feature maps through the initial convolution layer and max pooling layer;
3) Pass feature maps through stacked improved shuffle blocks embedded with SE modules, completing lightweight feature extraction with adaptive channel weighting;
4) Send the final feature map into global average pooling and full connection layers to output expression classification probabilities;
5) Use Focal Loss to calculate training loss and backpropagate gradients to update network parameters.

The improved structure maintains the original low parameter and low FLOPs advantages of ShuffleNet V2, without introducing excessive computational overhead.

3.2 SE Attention Module Integration

The Squeeze-and-Excitation module realizes adaptive feature channel weighting through two core stages: Squeeze and Excitation.

1) **Squeeze Operation**: Apply global average pooling on each channel of the input feature map $\mathrm{X} \in \mathrm{R}^{\mathrm{H} \times \mathrm{W} \times \mathrm{C}}$, compressing spatial dimension information into a channel descriptor

$$Z_c\text{:}\mathrm{z_c} = \frac{1}{\mathrm{H} \times \mathrm{W}} \sum_{\mathrm{i}=1}^{\mathrm{H}} \sum_{\mathrm{j}=1}^{\mathrm{W}} \quad \mathrm{X_c}(\mathrm{i}, \mathrm{j}))$$

where H, W, C denote the height, width and channel number of feature maps respectively.

2) **Excitation Operation**: Two fully connected layers are used to learn nonlinear channel correlation. The first FC layer reduces channel dimensions with ReLU activation, the second FC layer recovers original channel number and maps weights to the interval [0,1] via Sigmoid function to obtain channel weight vector$s = \sigma(W2\delta(W1z))$ δ represents ReLU activation, σ represents Sigmoid activation.

3) **Feature Recalibration**: Multiply the original feature map channel by the corresponding weight $\mathrm{s_c}$ to obtain weighted feature output $\tilde{X}\text{:}\tilde{X}_c = s_c \cdot X_c$.

In this paper, SE modules are embedded after the channel shuffle operation of every ShuffleNet V2 block. The model automatically assigns larger weights to channels containing facial key areas such as eyes, mouth and eyebrows, suppressing useless background feature responses and strengthening the capture of subtle expression differences.

3.3 Multi-Dimensional Compound Data Augmentation

To improve model robustness against illumination change, shooting angle offset and noise interference in real scenarios, three types of augmentation strategies are combined to expand training samples:

1) **Geometric Transformation**: Random horizontal flip, slight rotation within $\pm 15°$, random scaling and random cropping, simulating different face shooting poses;
2) **Color Jitter**: Random adjustment of image brightness, contrast and saturation, adapting to complex indoor and outdoor lighting conditions;
3) **Noise Injection**: Add low-intensity Gaussian noise to training images to resist sensor noise interference of edge cameras.

All augmentation operations are only applied to the training set, while validation and test sets retain original images to ensure objective evaluation of model generalization. The original dataset is split into training set (70%), validation set (15%) and test set (15%) by stratified sampling to keep the original expression class distribution ratio.

3.4 Focal Loss for Class Imbalance Mitigation

Standard cross-entropy loss treats all samples equally. For FER datasets with unbalanced categories, the model tends to fit majority samples such as happy and neutral, ignoring minority samples like disgust and fear. Focal Loss introduces modulating factor $(1-p_t)^\gamma$ to reduce the loss weight of easy-classified samples and focus training on hard minority samples. The mathematical formula is defined as:

$$FL(p_t) = -\alpha_t (1-p_t)^\gamma \log(p_t)$$

Where p_t is the model predicted probability of the true category, α_t is the class balance weight coefficient, $\gamma \geqslant 0$ is the focusing hyperparameter. When $\gamma > 0$, the loss contribution of samples with high p_t (easy samples) is suppressed. In this experiment, we set $\gamma = 2$ and set α_t inversely proportional to the sample quantity of each expression category to balance class loss weights.

4. Experiments and Analysis

4.1 Experimental Setup

4.1.1 Datasets

Three public standard facial expression datasets are adopted for evaluation:

1) **FER2013**: Contains 35,887 grayscale face images divided into seven expression categories (angry, disgust, fear, happy, sad, surprise, neutral), with obvious class imbalance;
2) **CK+ (Cohn-Kanade Plus)**: Standard laboratory facial expression dataset with clear facial texture, covering six basic expressions proposed by Ekman;
3) **AffectNet**: Large-scale real-scene color expression dataset with diverse backgrounds, simulating complex application environments.

All images are unified to 48×48 resolution, normalized to \([0,1]\) and converted to three-channel RGB input for network training.

4.1.2 Hardware and Software Environment

- Hardware: NVIDIA RTX 3060 GPU with 12GB video memory;
- Software: Python 3.8, PyTorch 1.10, OpenCV, Albumentations data augmentation library;
- Training hyperparameters: Batch size = 32, initial learning rate = 0.001, Adam optimizer, maximum epoch = 100, early stopping strategy with patience=10 to prevent overfitting.

4.1.3 Evaluation Metrics

Four mainstream classification evaluation indicators are used: Overall Accuracy (Acc), Mean Recall (Recall), Mean F1-Score and model parameter quantity (Params) to comprehensively compare recognition performance and lightweight degree. Confusion matrix is drawn to analyze the recognition effect of each minority expression category.

4.2 Comparative Experiment Results

Table 1. Performance comparison of different lightweight models

Model	FER2013 Acc(%)	CK+ Acc(%)	AffectNet Acc(%)	Params(M)
ShuffleNet V2	63.21	89.57	58.74	1.26
MobileNet V2	64.75	90.13	60.12	2.28
ShuffleNet V2+SE	66.83	92.46	62.39	1.31
Proposed Model (SE+Aug+Focal Loss)	69.47	94.82	65.71	1.31

From Table 1, the proposed model achieves the highest recognition accuracy on all three datasets, with only a tiny parameter increment of 0.05M compared with vanilla ShuffleNet V2, fully retaining lightweight deployment advantages. On FER2013 with severe category imbalance, the accuracy rises by 6.26% compared with the original network, which verifies the effectiveness of Focal Loss in solving unbalanced sample bias. The model outperforms MobileNet V2 with far fewer parameters, proving its superiority for edge terminal deployment.

4.3 Ablation Experiment

Table 2. Ablation study results on FER2013

Setting	Accuracy(%)	Mean F1
Baseline (ShuffleNet V2)	63.21	0.603
Baseline + SE Attention	66.83	0.647
Baseline + SE + Data Augmentation	68.15	0.669
Baseline + SE + Aug + Focal Loss (Ours)	69.47	0.692

1) Adding SE attention alone brings a 3.62% accuracy improvement, demonstrating that the attention module effectively extracts discriminative facial emotional features;

2) Combining multi-dimensional data augmentation further improves accuracy by 1.32%, proving that expanded diverse samples enhance model anti-interference ability;

3) Introducing Focal Loss increases accuracy by another 1.32%, significantly lifting the F1 value of minority categories such as disgust and fear, effectively alleviating dataset imbalance issues.

4.4 Visualization Analysis

The class confusion matrix of the proposed model on FER2013 shows that the recognition recall rate of minority expressions disgust and fear is greatly improved compared with original ShuffleNet V2, which is attributed to the focusing mechanism of Focal Loss. Attention heatmap visualization displays that the model concentrates feature responses on facial emotional key areas including eyes, eyebrows and mouth, rather than irrelevant background regions, proving the SE module's feature screening capability.

5. Conclusion and Future Work

Aiming at the problems of low feature discrimination and category imbalance in lightweight facial expression recognition based on ShuffleNet, this paper proposes an improved ShuffleNet model embedded with SE attention, compound data augmentation and Focal Loss. SE attention modules are inserted into shuffle blocks to adaptively weight expression-sensitive feature channels; multi-dimensional augmentation strategies expand training sample diversity to strengthen environmental robustness; Focal Loss suppresses the loss weight of easy majority samples and improves the recognition precision of rare facial expressions. Experimental results on FER2013, CK+ and AffectNet datasets indicate that the proposed method achieves higher recognition accuracy than original ShuffleNet V2 and MobileNet V2 under ultra-light parameter scale, satisfying the real-time deployment demand of facial expression detection on mobile and embedded devices.

In future research, we will carry out two aspects of optimization:

1) Integrate lightweight coordinate attention (CA) to replace partial SE modules to capture spatial position information of facial organs;

2) Combine model quantization and network pruning technology to further compress model volume and reduce inference latency

for low-power embedded chips.

References

Ekman P, Friesen W V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124-129.

Lai Z Y, Chen R H, Qian Y R. (2020). Real-time micro-expression recognition algorithm based on atrous convolutions for CNN. *Application Research of Computers, 37*(12), 3777-3780, 3835.

Li D H, Zhong T, Wang S, et al. (2024). Lightweight identification of tomato leaf diseases based on improved ShuffleNet v2. *Jiangsu Agricultural Sciences, 52*(03), 220-228.

Li F. (2020). Research on facial expression recognition technology based on deep learning. Dalian Maritime University.

Peng K B, Lv X J, Li C, et al. (2024). Research on railway passenger station security risk assessment technology based on contraband detection and facial expression recognition. *Railway Transport and Economy, 46*(01), 109-115.

Qu F, Wang S J, Yan W J, et al. (2018). CAS(ME)^2: A Database for Spontaneous Macro-Expression and Micro-Expression Spotting and Recognition. *IEEE Transactions on Affective Computing, 9*(4), 424-436.

Rong Y, Liu J W, Li X L. (2024). Adaptive hybrid network for students' classroom affective computing. *Journal of Computer Applications, 29*(6), 1607-1627.

Tao J H, Fan C H, Lian Z, et al. (2024). Development status and trend of multimodal emotion recognition and understanding. *Journal of Image and Graphics, 29*(06), 1607-1627.

Wang X R, Zhang Y T, Peng X. (2022). Research on vehicle fatigue driving system based on face key point detection. *Wireless Internet Technology, 19*(17), 82-84.

Yang Z, Wu S D, Jia R. (2024). Autonomous Classification and Early Warning Framework for UAV Images Combining Improved ShuffleNet-V2 and Attention Mechanism. *Journal of Wuhan Institute of Chemical Technology, 47*(5), 89-95.

Zhou Y, Chen S C, Wang Y M, et al. (2020). Review of research on lightweight convolutional neural networks. 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference. *IEEE*, 1713-1720.

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.paradigmpress.org/jpeps
Volume 5 Number 2 June 2026

The Four Necessary and Sufficient Canonical Tests of the Special Theory of Relativity

Adrian Sfarti[1]

[1] UC Berkeley, CS Dept, 387 Soda Hall, Berkeley, CA, USA
Correspondence: Adrian Sfarti, UC Berkeley, CS Dept, 387 Soda Hall, Berkeley, CA, USA.

doi:10.63593/JPEPS.2026.06.03

Abstract

The present paper is a continuation of our previous published work where we dispelled the misconception that the three canonical experiments mentioned by H.P. Robertson in his 1949 seminal paper, are all that is necessary and sufficient in deciding if and when an alternative theory of relativity is equivalent to Einstein's special relativity.

Keywords: special relativity, Selleri transformations, Lorentz transformations, relativistic Doppler effect, relativistic optical aberration, Fizeau experiment, Michelson-Morley experiment in non-vacuum medium, Robertson 1949 paper

1. Introduction

In our previously published paper (Sfarti, A., 2025), we have shown that one produces contradictory outcomes by inserting the Seller transforms versus the Lorentz transforms into the plane wave formula. We remind you that inserting the Lorentz (Einstein, A., 1905; Lorentz, H. A., 1904) transformations into the plane wave equation Einstein obtained through identification of the coefficients multiplying x, y, z, t respectively:

$$\omega = \omega' \gamma (1 + \frac{l'v}{c})$$
$$l = \frac{l' - \frac{v}{c}}{1 + \frac{l'v}{c}} \tag{1.1}$$

On the other hand, inserting the Selleri transformation the plane wave equation one obtains:

$$\omega = \omega' \gamma(v)(1 + \frac{l'v}{c})$$
$$l = \frac{l'}{1 + \frac{l'v}{c}} \tag{1.2}$$

The Selleri derivation of the relativistic Doppler effect is correct but the relativistic aberration formula is not correct as we can see by comparison with Einstein, and, more importantly, by comparison with experiment. There is an obvious difference in the cosine for the x direction while the predictions for the y and z directions coincide between the two theories. Today, we know that Selleri transformation would pass the Ives-Stilwell experiment that tests the transverse Doppler effect (Saathoff, G. et al., 2003; Müller, H. et al., 2003a; 2003b; 2003c). The above derivation shows that the Selleri theory would give an incorrect prediction for all optical experiments involving relativistic aberration. In addition, Selleri's transformations give the wrong answer for the speed composition formula:

$$u_x^{'} = \frac{dx'}{dt'} = \frac{dx - vdt}{dt} = u_x - v \tag{1.3}$$

$$u_y^{'} = \frac{dy'}{dt'} = \frac{dy}{\gamma dt} = \frac{u_y}{\gamma} \tag{1.4}$$

The expression (1.3) above means that the Selleri theory gives the wrong prediction for the Fizeau experiment while (1.3) and (1.4) combined show that the Selleri theory gives the wrong prediction for all Michelson-Morley experiments in refractive (non-vacuum) media (Sfarti, A., 2011). According to H.P. Robertson's seminal paper (Robertson, H.P., 1949), Selleri's theory should be equivalent with the special theory of relativity since it gives the same answers for the canonical tests of special relativity, the Michelson-Morley, Kennedy-Thorndike and Ives-Stilwell experiments (Saathoff, G. et al., 2003; Müller, H. et al., 2003a; 2003b; 2003c; Sfarti, A., 2011; Sheng, S., 2024; The KM3NeT Collaboration, 2025; Jin, T., 2024; Wei, J. J., 2025; Wolf, W., 2024; Kepuladze, Z., 2025; Amram, D., 2024; Brout, D., 2022; Guerrero, M., 2025; Li, J.R., 2025). Yet, we see that the answers diverge when it comes to the tests of relativistic aberration. This brings us to the realization that the three canonical optical tests are necessary but not sufficient in establishing the equivalence between the special theory of relativity and alternative interpretations, tests of relativistic aberration need to be added to the set. Alternatively, tests of the Michelson-Morley in refractive medium should be added. This conclusion follows naturally from the fact that the Michelson-Morley experiment deals with space transformations, the Ives-Stilwell experiment is an expression of time transformation (time dilation) while the Kennedy-Thorndike experiment is an expression of both time and space transformations while aberration is an expression of angle transformation, so it is not covered by any of the three tests mentioned by Robertson and this is how he missed the issue.

2. Relationships with the Test Theories of Special Relativity

There are two main types of tests of special relativity: the Robertson-Mansouri-Sexl (Mansouri R. & Sexl R.U., 1977) (in short, RMS) and the Standard Model Extension (Kostelecký, V. A. & Russell, N. 2011) (in short, SME). Both theories are parametrized theories of special relativity (SME including RMS by virtue of having much more parameters). For example, RMS has three parameters, α, β, δ that relate respectively to tests of time dilation, measurement of length in the direction of motion and measuring length in the direction perpendicular on the direction of motion. The combination of the three canonical experiments is necessary to obtain the complete Lorentz transformation. Michelson–Morley only tests the combination between β and δ, while Kennedy–Thorndike tests the combination between α and β. To obtain the individual values, it's necessary to measure one of these quantities directly. This was achieved by Ives–Stilwell experiment that measured α. So β can be determined using Kennedy–Thorndike, and subsequently δ using Michelson–Morley. Deviations from the two-way (round-trip) speed of light are given by:

$$\frac{c}{c'} = 1 + (\beta - \delta - 0.5)\frac{v^2}{c^2}\sin^2\theta + (\alpha - \beta + 1)\frac{v^2}{c^2} \tag{2.1}$$

where c is the speed of light in the "preferred" frame, and c′ is the speed of light measured in the moving frame at an angle θ from the direction in which the frame is moving. To verify that special relativity is correct, the expected values of the parameters are $\alpha = -0.5, \beta = 0.5, \delta = 0$ and thus c′=c (Sfarti, A., 2008a; 2008b; 2014).

In the previous section we have shown that a fourth parameter, say φ, is necessary to be associated with a fourth canonical test, the one that measures angle. In other words, the current three canonical experimental tests of special relativity are necessary but not sufficient, a fourth test, relating to measuring aberration angles is necessary. An example as how the RMS theory could be extended via the addition of the extra parameter is by adding a parametrized angle transformation:

$$l = \frac{l' - \varphi \frac{v}{c}}{1 + \frac{l' v}{c}} \tag{2.2}$$

Using the notation from (2.1) the above can be re-written as:

$$\cos\theta = \frac{\cos\theta' - \frac{v}{c}\varphi}{1 + \frac{v}{c}\cos\theta'} \qquad (2.3)$$

For $\varphi = 1$ one recovers special relativity. Note that we need that **both** the speed **and** the angle of light ray constrain **both** the parameters α, β, δ **and** φ. The constraint of parameter φ is by far the strongest due to the dependency in first order of the speed v.

3. Important Contemporary Reviews of the Selleri Relativity

An excellent paper (Capria, M., 2022) by Marco Capria does an in-depth analysis of Selleri's "Weak Relativity" concluding with the previous observations that while Selleri theory is "experimentally indistinguishable" from special relativity, its explanations of certain experiments, most notably of the Sagnac experiment, are downright incorrect, bordering on fringe, in the style of Ives or Dingle (see his paper conclusions). Like everybody before him, Capria follows the path set by H.P. Robertson (Robertson, H.P., 1949) in considering only the Michelson-Morley, Kennedy-Thorndike and Ives-Stilwell experiments as canonical experiments in determining equivalence between various theories and the special theory of relativity, thus missing the fact that experiments in the class of Fizeau, stellar aberration or Michelson-Morley in non-vacuum media provide the "separation" between Selleri's and Einstein's relativity.

4. Conclusion

We have discovered that the three canonical optical tests are necessary but not sufficient in establishing the equivalence between the special theory of relativity and alternative interpretations, the tests of relativistic aberration need to be added to the list.

References

Amram, D. (2024). New Constraint for Isotropic Lorentz Violation from LHC Data. *Phys. Rev. Lett., 132*, 21180. DOI: https://doi.org/10.1103/PhysRevLett.132.211801

Brout, D. (2022). The Pantheon + Analysis: Cosmological Constraints. *The Astrophysical Journal, 938*(2).

Capria, M. (2022). On Selleri's "Weak Relativity". https://arxiv.org/abs/2209.09551

Einstein, A. (1905). On the Electrodynamics of Moving Bodies. *Ann. der Phys.*

Guerrero, M. (2025). Bounding anisotropic Lorentz invariance violation from measurements of the effective energy scale of quantum gravity. *Physical Review D.*, 112. DOI: https://doi.org/10.1103/k3xg-wkrc

Jin, T. (2024). Constraint of d = 8 Lorentz Invariance Violation with New Torsion Pendulum Experiment. *Symmetry, 16*(11), 1432. https://doi.org/10.3390/sym16111432

Kepuladze, Z. (2025). Lorentz Violation: Loop-Induced Effects in QED and Observational Constraints. arxiv.org/abs/2504.15608

Kostelecký, V. A., Russell, N. (2011). Data tables for Lorentz and CPT violation. *Reviews of Modern Physics, 83*(1), 11–32. arXiv:0801.0287. Bibcode:2011RvMP...83...11K. doi:10.1103/RevModPhys.83.11

Li, J.R. (2025). Experimental design for testing local Lorentz invariance violation in gravity. *Class. Quant. Grav., 42*(2), 025012.

Lorentz, H. A. (1904). Electromagnetic Phenomena in a System Moving with Any Velocity Smaller than That of Light. *Proceedings of the Royal Netherlands Academy of Arts and Sciences, 6*, 809-831.

Mansouri R., Sexl R.U. (1977). A test theory of special relativity: III. Second-order tests. *Gen. Rel. Gravit., 8*(10), 809–814. doi:10.1007/BF00759585. S2CID121834946

Müller, H., Braxmaier, C., Herrmann, S., Peters, A., Lämmerzahl, C. (2003c). Electromagnetic cavities and Lorentz invariance violation. *Phys. Rev. D, 67*.

Müller, H., Herrmann, S., Braxmaier, C., Schiller, S., Peters, A. (2003a). Modern Michelson-Morley experiment using cryogenic optical resonators. *Phys. Rev. Lett.*, 91.

Müller, H., Herrmann, S., Braxmaier, C., Schiller, S., Peters, A. (2003b). Theory and technology for a modern Michelson-Morley Test of Special Relativity. *App. Phys. B*.

Robertson, H. P. (1949). Postulate versus Observation in the Special Theory of Relativity. *Rev. Mod.Phys., 21*(3).

Saathoff, G., Karpuk, S., Eisenbarth, U., Huber, G., Krohn, S., Horta, R/M., Reinhardt, S., Schwalm, D., Wolf, A., Gwinner, G. (2003). Improved test of time dilation in special relativity. *Phys. Rev. Lett., 91*(19).

Sfarti, A. (2008a). The Mansouri-Sexl formalism for the Michelson-Morley experiment in non-vacuum media. *Can. J. Phys,* 86.

Sfarti, A. (2008b). The Mansouri-Sexl formalism for the Sagnac experiment. *Can. J. Phys, 86.*

Sfarti, A. (2011). Comment on the Paper "Absolute Motion Determined from Michelson-Type Experiments in Optical Media" by V. P. Dmitriyev. *Zeitschrift für Naturforschung, 66a*.

Sfarti, A. (2014). Tests of Light Speed Isotropy: Experimental Method for Constraining the Mansouri-Sexl Parameter α by Using Ring Laser Gyroscopes. *Open J. Mod. Phys, 1*(2).

Sfarti, A. (2025). Selleri Theory Not Equivalent to the Special Theory of Relativity and a Related Note on H.P. Robertson 1949 Note. *JPEPS, 4*(6).

Sheng, S. (2024). Precise spectroscopy of metastable Li+ using the optical frequency comb in support of time dilation tests. *Photonics Research, 13*(1). DOI:10.1364/PRJ.538659

The KM3NeT Collaboration. (2025). KM3NeT constraint on Lorentz-violating superluminal neutrino velocities. *Comm. Phys, 8*(1), 457.

Wei, J. J. (2025). New Tests on Lorentz Invariance Violation Using Energy-resolved Polarimetry of Gamma-Ray Bursts. *Astrophysical Journal, 983*(2).

Wolf, W. (2024). Underdetermination in Classic and Modern Tests of General Relativity. *Philosophy of the Natural Sciences, 14*(57).

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.paradigmpress.org/jpeps
Volume 5 Number 2 June 2026

Anthropometric Studies of the Diurnal Human Height Shrinkage Across Selected Professionals in Nigeria

Mojisola A. Bolarinwa[1] & Charles O. E. Owaba[1]

[1] Department of Industrial and Production Engineering, University of Ibadan, Nigeria
Correspondence: Mojisola A. Bolarinwa, Department of Industrial and Production Engineering, University of Ibadan, Nigeria.

doi:10.63593/JPEPS.2026.06.04

Abstract

Diurnal variation in stature is one potentially significant source of error in the evaluation of short-term growth. Hence, there is the need to observe whether there are diurnal variations when assessing humans' (subjects) body composition, using physical anthropometric methods. In this paper, the diurnal human height shrinkage have been investigated by measurements, such that twice-daily measurements and body weight of subjects were examined and recorded. A total of 70 subjects were examined altogether. This comprises of 35 males and 35 females of which 2 nursery school children (a boy and a girl) were involved, aging 4 years and 3½ years respectively. For most of the subjects, Morning measurements were made immediately after rising. The time of the measurements was also recorded and diurnal changes during daytime were estimated. Results showed that stature decreased during the day in the entire 70 subjects surveyed and sampled. From the results, it could be seen that an average female subject will shrink by 0.88% of the usual (morning) height. Also, the mean average shrinkage in height amounts to 1.43 ± 2cm for the females. Likewise, an average male subject will shrink by 0.91% of his usual morning height. The mean average shrinkage in height in male subjects equals 1.56 ± 2cm. The findings of this study confirm the existence of diurnal variation and that the greater proportion of the height loss occurs during the earlier part of the day, especially when subjects are engaged with some form of activity or the other.

Keywords: variations, short term growth, stature, measurements, shrinkage

1. Introduction

The importance of minimizing measurement error in the assessment of growth has been well documented. (Voss et al., 1990; Voss et al., 1991; Voss & Bailey, 1994; Wasse, 1726). One potentially significant source of error, diurnal variation in stature, first noted in 1724 (Redfield & Meredith, 1938) has, however, been largely ignored in clinical practice. Early studies by Redfield and Meredith (Boyd, 1929; Nylin, 1929) were conducted with varying degrees of scientific rigour, but did confirm the presence of diurnal variation in the adult. Most agreed that the total loss amounted to between 2 and 3 cm, and the evidence suggested that the greater proportion of the decrease in height was occurring in the trunk.

Similar effects have been shown in children, (Palmer, 1930; Strickland & Shearin, 1972; Whitehouse et.al., 1974; Buckler, 1978; Ashizawa & Kawabata, 1990; LampI, 1992), some studies also showing that much of the height loss can be restored by taking a short nap (Nylin, 1929; Palmer, 1930; Strickland & Shearin, 1972; Whitehouse et.al., 1974; Buckler, 1978; Ashizawa & Kawabata, 1990; Lampi, 1992; Hrdlicka, 1939). Almost all reports agree that the greater proportion of the decrease in stature appears to occur soon after rising (Boyd, 1929; Strickland & Shearin, 1972; Ashizawa & Kawabata, 1990) though it is assumed that, without a nap, further loss continues throughout the day. Foundational work of Reilly et.al. (1984), established on investigating the circadian rhythm of stature in eight adult males using precision stadiometry, reported a trough-to-peak variation of 19.3 mm, approximately 1.1% of total stature. Krishan and Vij (2007) measured four adults and one child over 56 days and reported a maximum mean daytime stature loss of up to 28.1 mm, with a very rapid decrease in the first two hours after rising. Wilby et al. (1987) documented a mean peak-to-trough variation of 15.4 mm (0.92% of stature) in ten females, somewhat less than the ~19 mm reported in males. In children, Lampl (1992) measured one adolescent boy over 300 days and found a mean daily decrease of 9.8 ± 2.0 mm, with similar decreases observed after short naps. Tillmann and Clayton (2001) studied two prepubertal boys and reported a maximal height loss of about 14.4 mm by 15:00 hours, followed by a significant partial recovery of approximately 3.8–4.2 mm during the evening hours suggesting the diurnal pattern is not strictly monotonic. There is some disagreement about the total daily loss to be expected, but no two studies have measured their subjects over exactly the same period. Even in two studies involving large number of children, one found a mean decrease in height of 1.54 cm in 100 children between rising and late afternoon, whereas another found a mean decrease of just 1.0 cm in 70 boys between early morning and bedtime (Palmer, 1930; Whitehouse et.al., 1974). In their own findings, Czaprowski et al. (2019) confirmed diurnal shrinkage in 98 children with idiopathic scoliosis, noting a mean standing height loss of 7 mm from morning to evening. A study of 197 Indian medical students demonstrated significant diurnal variation between morning (08:00–09:00) and afternoon (16:00–17:00) measurements, with males showing greater height loss than females (Rajesh et al., 2022). The underlying mechanism surrounding diurnal stature change is the viscoelastic creep behavior of the intervertebral discs (IVDs). Botsford, et. al. (1994), using three-dimensional MRI reconstructions quantified diurnal disc volume changes in vivo, reporting mean decreases of 18–22% in lumbar disc volume after a simulated day of activity. Keller and Nathan (1999) developed a sagittal-plane viscoelastic model predicting an instantaneous height loss of 11.7 mm upon loading and a total of 19.6 mm after eight hours, with the thoracic region contributing about 57% and the lumbar region about 32% of total loss due to larger number of thoracic discs. Martin et al. (2022), using quantitative MRI were able to show that lumbar discs deform by 6–8% in height and 5–9% in volume from morning to afternoon, with deformations concentrated posteriorly at caudal levels; notably at L5-S1, where disc herniation is most prevalent. Vergroesen et al. (2016) provided a new biophysical explanation, stating that as in vivo daytime loading is sustained for roughly twice as long as nighttime unloading, the disc operates closer to its loading equilibrium during the day. Zander et al. (2010) used finite element modeling to demonstrate that a 10% diurnal change in disc height produces markedly higher spine flexibility and facet joint forces in the evening, as compared to the morning, while intradiscal pressures are higher in the morning; a finding with implications for injury risk at different times of day. Adams et al. (1990) reviewed these diurnal mechanical changes and reported that different spinal structures bear disproportionate loads at different times of the day, and that the timing of symptom onset may help clinicians understand the pathophysiology of back pain. Factors responsible for the modulation of the magnitude of diurnal shrinkage have been found to include: Body weight and obesity; External loading and activity type (Tyrrell et al., 1985; Fowler et al., 1994; Hoe et al., 1994; Van Deursen et al., 2005; Jazwinska & Adam, 1985; Reilly & Freeman, 2006); as well as low back pain (Healey et al., 2008; Healey et al., 2011). Precision stadiometry is the standard technique for quantifying stature change. Eklund and Corlett (1984) developed an early device capable of measuring height with a standard deviation below 1 mm and demonstrated its utility for ergonomic evaluations of workplace loading. Foreman and Linge (1989) identified heel pad compression as a confounding factor, reporting an average of 4.4 mm of heel compression upon standing, and recommended a two-minute standing equilibration period before measurement. Towards the end of the last century, attempts were made largely by the physical anthropologists to standardize the method of measurement (Thomsen et al.,1990). These authors claim to have shown that, using this technique, loss in stature between morning and afternoon, though not entirely eliminated, can be reduced to a maximum of 0.46 cm. However, the technique has changed little over the years, with clinicians showing little interest in the subject. For instance, stretching technique was not widely adopted about 20 years ago, until discovered could minimize the effects of diurnal variation (Buckler, 1978). In his own case, Thomsen et al (1990) compared the precision or reproducibility of the stretched and un-stretched methods and report no difference. In the recent times, measurements of reduction in stature have been used to compare spinal loading in chronic low-back pain (CLBP) and asymptomatic populations, even though whether there are any differences in the repeatability of stature measurements, between those with and without CLBP, is not known (Healy et al., 2005). According to Tyrell et al. (1985), using a method comparable with that of Eklund and Corlett (1984), stature was measured with an accuracy of 1 mm in eight young adults. Fifty-four percent of the diurnal loss in stature occurred in the first hour after rising. Approximately 70% was regained during the first half of the night. Stothart and McGill (2000) showed that leaving subjects in place during repeated measures reduces random variation compared to having them step in and out, though it introduces a steady shrinkage artifact during the 3–3.5 minutes of measurement. Lewis and Fowler (2009) validated stadiometry against MRI, finding a significant correlation between stadiometric stature loss and posterior lumbar spine length change, supporting stadiometry as a reasonable proxy for overall spinal loading, though not for individual disc height measurement. Mikula et al. (2016) found that clinical height measurements in routine practice are highly unreliable, with 18% of patients showing recorded height differences of ≥2 cm over just three months in electronic health records, underscoring the need for standardized technique. Diurnal height variation has several practical consequences. In pediatric growth assessment, Tillmann and Clayton (2001) recommended measuring children between 18:00 and 21:00 when diurnal variation is smallest to minimize error in short-term growth evaluations. In forensic anthropology, failure to account for diurnal variation may introduce errors of 1–2 cm in stature estimation. In ergonomics, spinal shrinkage has been used as a noninvasive index of cumulative occupational loading. McGill et al. (1996) cautioned, however, that substantial within-subject and between-subject variability limits the precision of shrinkage measurements in workplace assessments. Massey et al. (2012) showed through finite element modeling that disc degeneration reduces osmotic pressure by approximately 75% and decreases diurnal fluid exchange by ~21%, potentially compromising nutrient delivery to the avascular disc. Previous estimates of the diurnal changes in disc height have used radiologic, stereo-photographic, and magnetic resonance imaging e techniques. In this study, diurnal human height shrinkage has been investigated manually using anthropometric measurements. Hence, the objectives of this work include: (1) Showing that there is usually diurnal shrinkage in human height everyday; (2) Determining in which sex does shrinkage occurs greater (or lesser); (3) Determining if the severity of shrinkage (assuming it occurs) varies with profession of the subjects in question.

2. Methods

A random sampling of subjects across different professions was conducted at different periods of the day (that is, in the morning and evening time. Professions (disciplines) of subjects covered include Barbing, Electrician, NEPA official, Motor Mechanic, Hair dresser, Students, Lecturers.

Nursery school pupils, Farmer, Carpenter, Fashion designer, and Bus conductor to mention a few. Having taken all necessary precautions such as avoiding parallax errors by taking every reading from center points and on flat surfaces, the use of the non-shrinking type of the measuring tape for all measurements taken, removal of shoes and trinkets by subjects and so on, the human height (standing) was measured vertically from the floor to the top of the head (vertex with the aid of the measuring tape). Each person stood erect, looking ahead, the arms hanging loosely at the sides. Likewise for the weight, using the home scale with 1 kg calibration, corresponding values were recorded only when subjects must have assumed still positions, looking straight with shoulders kept erect with no movements at all. The stadiometer was used to note the exact height of subjects on the walls before marking corresponding points with pencils, in order to have accurate values.

During this activity, both the weights and corresponding heights of all subjects concerned were taken and recorded just after every subject rose from bed in the morning and later in the evening, between after the day's job and bedtime to verify if any noticeable change(s) will be obtained. In the end, the differences in the heights of the subjects in question were calculated to know if any reduction (shrinkage) has actually taken place. The time was of performance of every exercise were also noted and recorded. In all, about 70 subjects (35 females and 35 males) were visited, with readings taken from them at least twice in a day. Some of the subjects had to be visited repeatedly for days, with their weights and heights repeatedly checked as the outcomes obtained were equally surprising to them, and these they found difficult to accept easily.

The data obtained from this exercise are as shown in Table 2. Detailed analysis (descriptive and inferential statistical), results and discussions are as given under section 3. Table 1 shows some of the statistical formulae used for descriptive statistics. All readings were taken in centimeters for ease of understanding and conversion of data obtained.

2.1 Determination of Heights of the Subjects (Bolarinwa et, al., 2026)

1) Using the floor as a flat surface, each subject was in turn asked to remove shoes and caps, and thereafter asked to stand upright, on the stadiometer (base), positioned on the floor, with heels positioned closely together, arms by the sides, back as straight as possible and knees not bending in any way.

2) Before reading the height of each of the subjects, it was ascertained that: back of the head, the shoulders and buttocks all touch the vertical board of the stadiometer, with the eyes positioned in a straight, not tilted manner.

3) At this junction, the horizontal headpiece of the stadiometer was slowly brought down, and as such, it safely touched the top of the head of each subject in question.

4) Thereafter, the height was captured at the eye level to avoid parallax error, and recorded immediately. These procedures were repeated twice for each of the subjects involved in the research, that is, in the morning immediately on arrival in school, and in the afternoon, after the closing time.

Table 1. Key Statistical Formulae Used

Test	Formula
Paired t-test	$t = (\bar{x}\_diff) / (s\_diff/\sqrt{n})$
Welch's t-test	$t = (\bar{x}_1 - \bar{x}_2) / \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$
Mann-Whitney U	$U = n_1 n_2 + n_1(n_1+1)/2 - R_1$
Pearson r	$r = \Sigma[(x_i-\bar{x})(y_i-\bar{y})] / \sqrt{[\Sigma(x_i-\bar{x})^2 \times \Sigma(y_i-\bar{y})^2]}$
Cohen's d	$d = (\bar{x}_1 - \bar{x}_2) / s\_pooled$
One-way ANOVA	$F = (MS\_between) / (MS\_within)$

3. Data Analysis, Results and Discussions

Table 2. Data Captured from the Seventy Subjects Involved in the Study (Left)

Subject No	Gender	Age (Years/ Adult)	Occupations Discipline	Time 1 (am)	Weight 1(kg)
1.	F	62	Farmers wife	7:47	65

2.	F	30	Ind. Engrg. P.G. Student	4:30	56
3.	F	$3\overline{2}$	Nursery School Pupil	7:30	15.5
4.	F	Adult	P.G. Student	10:07	55
5.	F	Adult	P.G. Student	10:15	52
6.	F	Adult	P.G. Student	10:20	65
7.	F	Adult	P.G. Student	10:27	52
8.	F	Adult	P.G. Student	10:28	71
9.	F	Adult	P.G. Student	10:29	53
10.	F	Adult	P.G. Student	10:37	55.5
11.	F	Adult	P.G. Student	10:39	51
12.	F	Adult	P.G. Student	10:42	36.2
13.	F	Adult	P.G. Student	10:47	49
14.	F	Adult	P.G. Student	10:54	67.5
15.	F	Adult	P.G. Student	11:00	69
16.	F	Adult	P.G. Student	11:01	56
17.	F	Adult	P.G. Student	11:02	55.5
18.	F	Adult	P.G. Student	11:16	76
19.	F	Adult	P.G. Student	11:20	51
20.	F	Adult	P.G. Student	11:25	47
21.	F	Adult	P.G. Student	11:33	76
22.	F	Adult	P.G. Student	11:41	72
23.	F	Adult	P.G. Student	11:42	43
24.	F	Adult	P.G. Student	11:50	54
25.	F	Adult	P.G. Student	11:56	46
26.	F	Adult	P.G. Student	07:07	55
27.	F	Adult	P.G. Student	07:08	52.5
28.	F	Adult	P.G. Student	07:20	55
29.	F	Adult	P.G. Student	07:25	50
30.	F	Adult	P.G. Student	08:07	60.5
31.	F	Adult	P.G. Student	08:15	65
32.	F	Adult	P.G. Student	08:37	50
33.	F	Adult	P.G. Student	08:42	58.5
34.	F	Adult	P.G. Student	08:45	60
35.	F	Adult	Fashion designer	09:33	65
	Male				
36.	M	72	Farmer	07:45	56
37.	M	56	Professor	10:12	71
38.	M	38	NEPA Staff	08:10	65

39.	M	51	Motor Mech.	08:42	72
40.	M	26	Carpenter	08:47	63
41.	M	4	Nursery School Pupil	07:15	15
42.	M	Adult	Medical Stud	10:07	70
43.	M	Adult	P.G. Student	08:05	71
44.	M	Adult	P.G. Student	08:44	61
45.	M	Adult	P.G. Student	08:50	56
46.	M	Adult	P.G. Student	08:51	69
47.	M	Adult	P.G. Student	08:52	58.5
48.	M	Adult	P.G. Student	09:09	50
49.	M	Adult	P.G. Student	09:17	58
50.	M	Adult	P.G. Student	09:25	60
51.	M	Adult	P.G. Student	09:42	57
52.	M	Adult	P.G. Student	09:55	63
53.	M	Adult	P.G. Student	10:02	67
54.	M	Adult	P.G. Student	10:15	59
55.	M	Adult	P.G. Student	10:20	55
56.	M	Adult	P.G. Student	10:25	60.5
57.	M	Adult	P.G. Student	10:33	63
58.	M	Adult	P.G. Student	10:37	45
59.	M	Adult	P.G. Student	10:40	53.5
60.	M	Adult	P.G. Student	10:45	70
61.	M	Adult	P.G. Student	11:00	62
62.	M	Adult	P.G. Student	11:27	46
63.	M	Adult	P.G. Student	07:40	63
64.	M	Adult	P.G. Student	08:30	70
65.	M	Adult	P.G. Student	08:33	72
66.	M	Adult	P.G. Student	07:55	57
67.	M	Adult	P.G. Student	08:45	68
68.	M	Adult	Bus conductor	07:02	65
69.	M	Adult	Electrician	07:28	58
70.	M	Adult	Barber	08:05	60

Table 2. Data Captured from the Seventy Subjects Involved in the Study (Right)

Height 1(cm)	Time 2 (pm)	Weight 2 (kg)	Height 2 (cm)	Height2-Height 1 (cm)	Comments
162	4:30	65	161	1.0	Weight remained constant but height reduced
167.5	11:15	54.5	165	2.5	Weight and height

					reduced
100	08:30	16	98	2.0	Weight increased but height reduced
162.4	06:57	57	161.5	0.9	As in subject 3
162.5	08:31	52	160.5	2.0	As in subject 1
166	10:38	65	165	1.0	As in subject 5
153	07:08	52	152	1.0	As in subject 5
162.5	07:10	71.5	160	2.5	As in subject 3
171	08:25	52	169	2.0	As in subject 2
171	07:15	56.5	170	1.0	As in subject 3
153	07:17	51.5	152	1.0	As in subject 3
153	07:25	36	152	1.0	As in subject 2
165	10:25	49	163	2.0	As in subject 1
166	09:59	69	163	3.0	As in subject 3
162.8	07:40	68	162	0.8	As in subject 2
168	10:29	56	167.5	0.5	As in subject 1
162	08:20	54.5	161.5	0.5	As in subject 2
166	08:22	75.5	164.8	1.2	As in subject 2
156	08:55	51	154.7	1.3	As in subject 1
156.5	10:16	46	154.5	2.0	As in subject 2
166	10:13	75	164.5	1.5	As in subject 2
180	10:50	71.5	179	1.0	As in subject 2
160	11:00	45	159.4	0.6	As in subject 3
166	09:08	53	164	2.0	As in subject 2
171.5	09:19	46	170.8	0.7	As in subject 1
168	08:00	55	166.5	1.5	As in subject 1
167.4	08:05	52.5	166	1.4	As in subject 1
161.8	08:20	55	161	0.8	As in subject 1
163.8	08:25	50	162	1.8	As in subject 1
168	07:20	62.5	167	1,0	As in subject 3
169	07:00	65	167	2.0	As in subject 1
163	07:45	50	162.5	0.5	As in subject 1
169.4	10:25	58.5	166.5	2.9	As in subject 1
168	06:30	60	167.2	0.8	As in subject 1
165	05:35	64	162.8	2.2	As in subject 2
167	04:35	55	165.5	1.5	As in subject 2
178	05:45	71.5	177	1.0	As in subject 3
173	06:00	63	172	1.0	As in subject 2

176	05:25	70	174.2	1.8	As in subject 2
164	05:30	62	162	2.0	As in subject 2
103	03:35	15.5	100	3.0	As in subject 3
174.8	08:03	69	173	1.8	As in subject 2
171	07:10	72	170	1.0	As in subject 3
173	10:25	60	170	3.0	As in subject 2
177	10:15	57	175.4	1.6	As in subject 3
188	09:20	69	186.5	1.5	As in subject 1
169	09:14	59	168	1.0	As in subject 3
167.5	10:08	52	165.5	2.0	As in subject 3
184	10:10	59	182.5	1.5	As in subject 3
167.5	10:07	63	167	0.5	As in subject 3
182.5	09:25	59.5	180.5	2.0	As in subject 3
174	08:56	64	173	1.0	As in subject 3
178	09:07	67	176.5	1.5	As in subject 1
172.8	09:16	60	170	2.8	As in subject 3
165.5	09:22	55	163.2	2.3	As in subject 1
169	09:10	61	168.5	0.5	As in subject 3
182.5	07:20	62	182	0.5	As in subject 2
161	09:25	45	159	2.0	As in subject 1
169	10:40	55	168	1.0	As in subject 3
189	09:37	70	188	1.0	As in subject 1
175	09:42	62	173.6	1.4	As in subject 1
162	09:47	44	161.8	0.3	As in subject 2
170.5	09:50	60	169	1.5	As in subject 2
182.8	09:12	71	182	0.8	As in subject 3
184	09:25	71	182.8	1.2	As in subject 2
174.5	09:40	56	173	1.5	As in subject 2
184	07:15	68	182.8	1.2	As in subject 1
176	06:35	64	173.2	2.8	As in subject 2
165	07:00	58	162.8	2.2	As in subject 1
158	06:42	61	155.1	2.9	As in subject 3
6,007.9			5953.4	54.5	

3.1 Data Analysis

From the data shown in Table 2, subjects 1-35 represent the females selected for the exercise while subjects 36-70 represent the males selected. Although their weights fluctuated (where in some remained constant, some increased while some dropped slightly), all their heights (the 70 subjects put together) reduced to some extent, depending on the nature of profession of the subject in question. It was also shown that there usually is reduction in height daily per individual.

3.1.1 Females

The sum of the heights measured in the morning time = 5,693.1 cm.

The sum of the heights when measured in the evening time = 5.643.2 cm. The difference = height in the morning, H_1 minus height in the evening, H_2 = 5,693.1cm - 5,643.2cm = 49.9 cm.

Hence, shrinkage in the female subjects summed up to 49.9 cm.

Then the percentage shrinkage of the female subjects =

$$\frac{\text{Sum of shrinkage of females}}{\text{Sum of Morning heights of females}} \times 100$$

$$\frac{49.9}{5{,}693.1} \times 100 = 0.876\% = 0.88\%$$

This means that an average female subject will shrink by 0.88% of her usual (morning) height.

Moreover, the mean (average) shrinkage of the 35 female subjects

$$= \frac{\text{Sum of shrinkage of females}}{\text{The total number of female subjects}}$$

$$= \frac{49.9}{35} = 1.4257 = 1.43$$

3.1.2 Males

The sum of the heights measured in the morning = 6.007.9 cm.

The sum of the heights when measured in the evening time 5.953.4 cm.

The difference = Height in the morning, H_1 minus height in the evening, H_2 = (6,007.9 - 5,953.4) cm = 54.5 cm.

Hence, shrinkage in the male subjects summed up to 54.5 cm.

Then, the percentage shrinkage of the male subjects =

$$\frac{\text{Sum of shrinkage in males}}{\text{Sum of morning heights of male}} \times 100$$

$$\frac{54.5}{6{,}007.9} \times 100 = 0.907 = 0.91\%$$

This implies that an average male subject will normally shrink by 0.91% of his morning height.

Likewise, the mean (average) shrinkage of the 35 male subjects.

$$\frac{\textit{The sum of shrinkage in males}}{\textit{The total number of male subjects}}$$

$$= \frac{54.5}{35}$$

= 1.5571

= 1.56

3.2 Descriptive and Inferential Statistics of Diurnal Height Loss

3.2.1 Descriptive Statistics

Table 3 summarises key descriptive statistics for the height-loss variable by gender. All values are in centimetres (cm).

Table 3. Descriptive Statistics for Diurnal Height Loss by Gender

Statistic	Female (n=35)	Male (n=35)
Mean	1.4257 cm	1.5571 cm
Median	1.2000 cm	1.5000 cm
Mode	1.0 cm (f=8)	1.0 cm (f=6)
Std Deviation (s)	0.7022 cm	0.7663 cm
Variance (s^2)	0.4931 cm^2	0.5872 cm^2
Std Error of Mean	0.1187 cm	0.1295 cm
Coeff. of Variation (CV)	49.26%	49.67%
Minimum	0.5 cm	0.2 cm
Maximum	3.0 cm	3.0 cm
Range	2.5 cm	2.8 cm
Q1 (25th Percentile)	0.95 cm	1.00 cm
Q3 (75th Percentile)	2.00 cm	2.00 cm
IQR (Q3 − Q1)	1.05 cm	1.00 cm

Skewness	+0.5681 (positive skew)	+0.3570 (positive skew)
Excess Kurtosis	−0.7070 (platykurtic)	−0.6347 (platykurtic)
95% CI for Mean	(1.1845, 1.6669) cm	(1.2796, 1.8061) cm

From Table 3, a number of inferences can be made. Overall mean diurnal height shrinkage across all 70 subjects was approximately 1.50 cm (SD = 0.73 cm). Median is very close to the mean, indicating an approximately symmetric central tendency. Although the mild positive skewness (+0.47) indicates a slight rightward tail — a few subjects experienced notably larger losses. Both distributions show a mild positive (right) skew: most subjects lose between 0.5 – 2.0 cm, while a small minority exceeds 2.5cm. The modal value of 1.0 cm was the single most frequently recorded loss (14 occurrences out of 70, or 20%). Negative excess kurtosis (−0.65) confirms a platykurtic distribution, meaning height-loss values are more uniformly spread around the mean than a normal curve would suggest. This occurrence, in both groups confirm values are more spread than a normal curve, reflecting occupational variability. The coefficient of variation (CV) of approximately 49% across the groups reflect moderate-to-high relative variability, which is expected, given the diversity of occupational and activity profiles among subjects. This means that the high CV underscores that diurnal shrinkage is far from uniform across individuals, that is, activity intensity is the dominant driver.

3.2.1.1 Frequency Distribution of Height Loss

Table 4 below shows the frequency distribution of diurnal height losses grouped into 0.5 cm class intervals for all 70 subjects.

Table 4. Frequency Distribution of Diurnal Height Loss for All 70 Subjects

Class Interval (cm)	Frequency (n)	Relative Freq. (%)	Cumulative (%)
0.0 – 0.5	1	1.4%	1.4%
0.5 – 1.0	14	20.0%	21.4%
1.0 – 1.5	20	28.6%	50.0%
1.5 – 2.0	12	17.1%	67.1%
2.0 – 2.5	14	20.0%	87.1%
2.5 – 3.0	6	8.6%	95.7%
≥3.0	3	4.3%	100.0%

As shown in Table 4, exactly 50% of all subjects (35 out of 70) experienced a height loss in the range of 1.0 to 2.0 cm. The most populated single class interval was 1.0–1.5 cm (28.6%), confirming the modal tendency around 1.0 cm. The combined upper tail (≥2.5 cm) accounts for 12.9% of subjects, predominantly among those with physically demanding or prolonged upright occupations.

3.2.2 Inferential Statistical Tests

Test 1 — One-Sample t-Test (Is Diurnal Height Loss Real?): A one-sample t-test was conducted to determine whether the observed mean height loss (1.50 cm) is statistically significantly different from zero (i.e., whether any real shrinkage occurs at the population level). The null hypothesis H_0 was: "mean diurnal height loss = 0." The result was $t(69) = 16.96$, $p = 2.56 \times 10^{-26}$, which overwhelmingly rejects H_0. This provides extremely strong statistical evidence that diurnal height shrinkage is a genuine, non-random physiological phenomenon. The effect is not artefactual.

Test 2 — Independent Samples t-Test (Male vs. Female): A Welch's independent-samples t-test was used to compare mean height loss between males ($\bar{x} = 1.56$ cm) and females ($\bar{x} = 1.43$ cm). The result was $t(67.9) = 0.667$, $p = 0.507$. There is no statistically significant difference in diurnal height loss between males and females at the $\alpha = 0.05$ level. The Cohen's d effect size of 0.16 is negligible, confirming the practical equivalence of height loss across gender in this sample. That is, gender alone does not drive shrinkage magnitude, rather, occupational physical loading is a more decisive factor.

Test 3 — Shapiro–Wilk Normality Test: The Shapiro–Wilk test indicated that the female subgroup (W = 0.917, p = 0.012) and the combined dataset (W = 0.943, p = 0.003) deviate significantly from normality. The male subgroup, however, does not significantly depart from normality (W = 0.951, p = 0.123). This mild non-normality in the female and combined datasets is consistent with the positive skewness observed and validates the use of a robust non-parametric check (Test 4).

Test 4 — Mann–Whitney U Test (Non-parametric M vs. F): Given the non-normality in the female group, the non-parametric Mann–Whitney U test was applied as a confirmatory comparison. The result (U = 674.0, p = 0.470) corroborates the parametric t-test: there is no significant difference in height-loss rank distributions between males and females. Both tests are in agreement.

Test 5 — Pearson Correlation (Initial Height vs. Height Loss): A Pearson correlation was computed between each subject's initial recorded height (Height 1, in cm) and their difference in height value. The result was r = −0.108, p = 0.379, indicating no statistically significant linear relationship. Taller individuals do not experience significantly greater or smaller absolute diurnal height loss than shorter individuals in this dataset.

3.3 Results and Discussions

Data collected from this research shows that work activities which involved repeated impacts led to a loss of stature (shrinkage) in humans. For example, subject numbers 1 and 36 are farmers, even though husband and wife but subject number 36 shrank more than subject number 1 because he had to do more work than his wife, meaning that he experienced more impacts and shocks on his spine, leading to a greater shrinkage. Likewise, looking into the various disciplines found in the data very critically and thoroughly, one would see that the shrinkages were greatest in the professions requiring rigorous activities e.g. farming, motor-mechanic, carpentry, medicine, fashion designing to mention a few. Again, by comparing subject numbers 34 and 35, respectively hairdresser and fashion designer, subject number 35 experienced a greater shrinkage because of the nature of her job which involves series of body motions, sitting posture and movements, unlike subject number 34 who would only need to occupy her hands in order to get her job done. In addition, subject number 2, an industrial engineering student and researcher who conducted these activities also checked and included her own measurements. It could be seen from the data collated that the shrinkage in her height was greater than those of most other female subjects.

Moreover, Diurnal human height shrinkage is more pronounced in men than in women, the reason is due to the fact that men perform more rigorous tasks than women, thus can be said that diurnal human height shrinkage is directly proportional to the nature of jobs or tasks. Experiments carried out showed that there is diurnal shrinkage during a working day and rapid recovery when lying down, especially over the night. Furthermore, the findings from descriptive statistics strongly support the hypothesis that diurnal height reduction is a universal, statistically significant physiological phenomenon. The mean loss of approximately 1.50 cm is consistent with the established biomechanical understanding that intervertebral disc compression under gravitational load throughout the day reduces spinal column length. Disc fluid (nucleus pulposus) is gradually expelled under axial loading during waking hours and is restored through disc rehydration during recumbent sleep.

The absence of a statistically significant sex difference is noteworthy. While males in this sample are taller on average and have a slightly higher mean loss (1.54 vs. 1.43 cm), the difference is not meaningful at $\alpha = 0.05$. This is supported by both the parametric and non-parametric tests. The implication is that diurnal height loss appears to be independent of sex in mixed occupational populations.

Larger losses (≥2.5 cm) were observed in subjects engaged in more physically demanding or prolonged upright activities: Subject 14 (P.G. Student, 3.0 cm), Subject 41 (Nursery Pupil, age 4, 3.0 cm), Subject 44 (P.G. Student, 3.0 cm), Subject 70 (Barber, 2.9 cm), Subject 33 (P.G. Student, 2.9 cm), and Subject 68 (Bus Conductor, 2.8 cm). The 4-year-old child (Subject 41) exhibiting 3.0 cm loss likely reflects proportionally greater disc hydration changes typical of pediatric spinal anatomy and warrants separate age-stratified analysis.

The lack of correlation between initial height and height loss (r = −0.108, p = 0.379) suggests that absolute stature is not a predictor of diurnal loss magnitude in this sample.

4. Conclusions

In this research, the diurnal human height shrinkage has been investigated by measurements, such that twice-daily measurements and body weight of subjects were examined and recorded. A total of 70 subjects were examined altogether. This comprises of 35 males and 35 females of which 2 nursery school children (a boy and a girl) were involved, aging 4 years and 3½ years respectively. For most of the subjects, Morning measurements were made immediately after rising. The time of the measurements was also recorded and diurnal changes during daytime were estimated. Based on these, the following conclusions have been drawn:

Diurnal height loss is a real and significant phenomenon, with variations established to be more pronounced in males than females.

Gender does not significantly influence height loss, although males show a slightly greater absolute loss. Occupation-driven activity intensity, and not biological gender, is the primary determinant.

Diurnal decrease in height do certainly occur in human beings (all subjects,) and even children are not exceptions, that is, it also occurs even in small children with greater part of the shrinkage taking place during the earlier part of the day.

Profession significantly influences height loss, with children showing the greatest loss, and postgraduate students the least. Manual labour workers showed elevated loss.

No interaction exists between gender and profession, suggesting that occupational demands affect height loss similarly in both gender.

Initial height does not predict height loss, indicating that height loss is a consistent phenomenon independent of stature.

Finally, the study provides adequate evidence that diurnal height loss is an all-round phenomenon. The key determinant of height loss magnitude appears to be occupational physical demands, rather than gender or initial stature. These findings have important implications for understanding spinal mechanics, occupational health, assessment of growth and spinal health across the lifespan. Diurnal human height shrinkage is primarily and largely a factor of spinal shrinkage. This is because spinal shrinkage is recognised as an index of the compressive forces acting on the spine. This shrinkage is caused by viscoelastic creep from compression of motion segments. When the discs are unweighted (such as during sleep), this process is reversed.

5. Recommendations

Based on the findings of this research, the following recommendations are suggested:

Manual labour workers should have regular spinal decompression breaks, and constantly demonstrate proper lifting techniques.

Students should have regular changes, and at the same time, avoid prolonged sitting.

Children should be monitored regularly, especially in relation to their activity levels. Also, adequate rest periods should be enforced on them.

Future studies should explore the role of occupation, activity duration, BMI, and age as covariates, using multiple regression modelling. A larger, age-stratified sample with corrected data recording would substantially strengthen the conclusions.

Acknowledgements

The authors wish to acknowledge the support and contributions of those who played one role or the other towards the success of this work, especially Mr Femi Ogunfuye, Mr Kunle Olaniyi and Mr Jeremiah Amogu.

References

Adams MA, Dolan P, Hutton WC, Porter RW. (1990). Diurnal changes in spinal mechanics and their clinical significance. *J Bone Joint Surgery Br*., *72*(2), 266-270.

Ahmed, M. L., Yudkin, P. L., Macfarlane, J. A., McPherson, K., Dunger, D. B. (1990). Are measurements of height made by health visitors sufficiently accurate for routine screening of growth? *Archives of Diseases of Children*, *65*(12), 1345-1348.

Ashizawa K. and Kawabata M. (1990). Daily measurements of the heights of two children from June 1984 to May 1985. *Annals of Human Biology*, *17*, 437-443.

Bolarinwa, M., Muyiwa, O. and Onyeukwu, S. (2026). Effect of Backpack on the Spinal Health of Lower-Level High School Students in Nigeria using *Taguchi*. *European Journal of Applied Science, Engineering and Technology,* *4*(3), 211-224.

Botsford D., Esses S., and Ogilvie-Harris D. (1994). In vivo diurnal variation in intervertebral disc volume and morphology. *Spine*, *19*(8), 935-940.

Boyd E. (1929). The experimental error inherent in measuring the growing human body. *J. Phys Anthropol*, *13*, 390-432.

Buckler, J. M. H. (1978). Variations in height throughout the day. *Arch. Dis. Child*, *53*, 762.

Czaprowski D., Tyrakowski, M., Bloda, J., Waś, J., Dembińska, A., Ewertowska, P. and Kotwicki, T. (2019). Diurnal variation of body height in children with idiopathic scoliosis. *J. Back and Musculoskelet Rehab*, *32*(2), 291-296.

Eklund, J. and Corlett, N. (1984). Shrinkage as a measure of the effect of load on the spine. *Spine*, *9*(2), 189-94.

Foreman T. and Linge K. (1989). The importance of heel compression in the measurement of diurnal stature variation. *Applied Ergonomics*, *20*(4), 299-301.

Fowler N., Lees A. and Reilly T. (1994). Spinal shrinkage in unloaded and loaded drop-jumping. *Ergonomics,* 37(1), 133-139.

Garbutt, G., Boocock, M. G., Reilly, I. and Troup, J. D. (1990). Running speed and spinal shrinkage in runners with and without low back pain. *Med Sci Sports Exercises*, *22*(6), 769-772.

Healey E., Burden A., McEwan I. and Fowler N. (2008). Stature loss and recovery following a period of loading. *Clinical Biomechanics*, *23*(7), 871-877.

Healey E., Burden A., McEwan I. and Fowler N. (2011). Diurnal variation in stature: do those with chronic low-back pain differ? *Clinical Biomechanics, 26*(4), 331-336.

Healey, E. L., Fowler. E., Burden, A. M., Mc Ewan, I. M. (2005). *Repeatability of Stature Measurements in with without chronic low back pain*. Taylor and Francis Publishing, 1613-1622.

Hoe, A., Atha, J. and Murray-Leslie C. (1994). Stature loss from sustained gentle body loading. *Annals of Human Biology*, *21*(2), 171-178.

Hrdlicka, A. (1939). Practical anthropometry. Wistar Institute, Philadelphia.

Jazwinska E. and Adam K. (1985). Diurnal change in stature: effects of sleep deprivation. *Experientia, 41*(12), 1527-1528.

Keller T. and Nathan M. (1999). Height change caused by creep in intervertebral discs: a sagittal plane model. *J. Spinal Disorder*, *12*(4), 313-324.

Krishan K. and Vij K. (2007). Diurnal variation of stature in three adults and one child. *Anthropologist*, 9(2), 113-117.

Lampl, M. (1992). Further observations on diurnal variation in standing height. *Annals of Human Biology*, *19*, 87-90.

Lewis, S. and Fowler, N. (2009). Changes in intervertebral disk dimensions and stature change measurements. *Archives of Physical and Med. Rehab, 90*(11), 1913-1918.

Linux Information Project (2005). Human Variability: A Brief Introduction.

Martin, J., Oldweiler, A., Kosinski, A., Spritzer, C., Soher, B. and Erickson M. (2022). Lumbar intervertebral disc diurnal deformations and T2 and T1rho relaxation times. *Eur Spine J.*, *31*(6), 1410–1419.

Massey, C., Van-Donkelaar, C., Vresilovic, E., Zavaliangos, A. and Marcolongo. M. (2012). Effects of aging and degeneration on the human intervertebral disc during the diurnal cycle. *J Orthop Res.*, *30*(1), 122-128.

McGill, S., Van-Wijk, M., Axler, C. and Gletsu, M. (1996). Studies of spinal shrinkage to evaluate low-back loading in the workplace. *Ergonomics, 39*(1), 92-102.

Mikula, A., Mikula, A., Hetzel, S., Binkley, N. and Anderson P. (2016). Clinical height measurements are unreliable: a call for improvement. *Osteoporos Int.*, *27*(7), 2259-2264.

Nylin, G. (1929). Periodical variations in growth, standard metabolism, and oxygen capacity of the blood in school children. *Acta Med Scand*, *31*(suppl), 207.

Palmer, C. E. (1930). Diurnal variations of height and weight in the human body during growth. *Anat Rec*, *45*, 234-235.

Rajesh, T., Kiran, PP., Rao, PU and Suresh, K. (2022). Study of diurnal variation of stature among MBBS students. *J. Indian Acad. Forensic Med.*, *44*(3).

Redfield, J. E. and Meredith, H. V. (1938).

Changes in the stature and sitting height of preschool children in relation to rest in the recumbent position and activity following rest. *Child Dev*, *9*, 293-302.

Reilly, T, Tyrrell A, and Troup J. (1984). Circadian variation in human stature. *Chronobiology International*, *1*(2), 121-126.

Reilly, T. and Chana, D. (1994). Spinal shrinkage in fast bowling. *Ergonomic*s, *37*(1), 127-132.

Reilly, T. and Freeman K. (2006). Effects of loading on spinal shrinkage in males of different age groups. *Applied Ergonomics*, *37*(3), 305-310.

Stothart J. and McGill S. (2000). Stadiometry: on measurement technique to reduce variability in spine shrinkage measurement. *Clinical Biomechanics*, *15*(7), 546-548.

Strickland, A. and Shearin, R. (1972). Diurnal height variation in children. *Pediatrics*, 72(80), 1023-1025.

Thomsen, J., Evald, P., Skieller, V. and Bjork, A. (1990). A comparative study of two different one methods of measuring stature and the velocity growth in children and adults. *Eur J. Orthodont;* 12:166-173.013

Tillmann, V. and Clayton, P. (2001). Diurnal variation in height and the reliability of height measurements. *Annals of Human Biology*, *28*(1), 68–79.

Tyrell, A., Reilly, T. and Troup, J. (1985). Circadian variation in stature and the effects of spinal loading. *Spine,* 10(2), 161-164.

Van-Deursen L., Van-Deursen, D. Snijders, C. and Wilke, H. (2005). Relationship between everyday activities and spinal shrinkage. *Clinical Biomechanics*, *20*(5), 547-550.

Vergroesen P., van der Veen, A., Emanuel, K., van Dieën, J. and Smit, T. (2016). The poro-elastic behaviour of the intervertebral disc. *J. Biomech*, *49*(6), 857-863.

Voss, L. D. and Bailey, B. J. R. (1994). Equipping the community to measure children's height. the reliability of portable instruments. *Arch Dis Child*, *70*, 469-471.

Voss, L., Bailey, B., Cumming, K., Wilkin, T. and Betts, P. (1990). The reliability of height measurement. *Arch Dis Child*, *65*, 1340-1344.

Voss, L., Wilkin, T., Bailey, B. and Betts, P. (1991). The reliability of height and height velocity in the assessment of growth. The Wessex Growth Study. *Arch Dis Child*, *66*, 833-837.

Vuvor F. and Harrison O. (2017). A study of the diurnal height changes among adults in Ghana.

Wasse, J. (1726). *Part of a letter from the Reverend Mr Wasse, Rector of Aynho in Northamptonshire, to Dr Mead, concerning the difference in the height of a human body, between morning and night*. Philosophical Transactions, 87-88.

Whitehouse R., Tanner, J. and Healy, M. (1974). Diurnal variation in stature and sitting height in 12-14-year-old boys. *Annals of Human Biology*, *1*, 103-106.

Wilby J, Linge K, Reilly T, Troup J. (1987). Spinal shrinkage in females: circadian variation and the effects of circuit weight-training. *Ergonomics*, *30*(1), 47-54.

Zander T, T Zander, P Krishnakanth, G Bergmann, A Rohlmann. (2010). Diurnal variations in intervertebral disc height affect spine flexibility. *Computer Methods Biomech Biomed Eng, 13*(5), 625-631.

Journal of Progress in Engineering
and Physical Science
ISSN 2709-4006
www.paradigmpress.org/jpeps
Volume 5 Number 2 June 2026

Thermo-Mechano-Electromagnetic Multi-Physics Coupling Analysis, Unified-Datum Tolerance Coordination and Full-Lifecycle Experimental Verification of Triple-Functional Integrated Automotive Laminated Windshield Glass

Yan Jiang[1]

[1] Wanqian Qianxiao, Dalian 116221, China
Correspondence: Yan Jiang, Wanqian Qianxiao, Dalian 116221, China.

doi:10.63593/JPEPS.2026.06.05

Abstract

Intelligent new-energy vehicles increasingly require front windshield laminated glass to integrate electrochromic dimming, low-temperature rapid defrosting and an embedded digital-television (DTV) antenna. The superposition of multilayer polymer films, screen-printed silver conductive traces and embedded flat wiring harnesses, however, introduces thermo-mechano-electromagnetic coupling failure modes. Here, we propose three hierarchically optimized substructures—a stepped-gradient ceramic anti-corrosion shielding layer, a 0.2 mm-protrusion full-wrap harness bonding configuration and full-wrap silicone sealing for marginal un-melted polyvinyl butyral (PVB) zones—and couple them with an ISO 5459-conformant unified symmetric datum (PD-A) tolerance chain model. A fully coupled thermo-mechano-electromagnetic finite-element numerical model, validated through a four-level mesh-convergence study and a five-factor parametric sensitivity matrix, is developed to quantify interfacial stress evolution and electromagnetic interference attenuation. A coupon-grade multi-station test fixture (denoted 8S10Z-39J-900) is fabricated and used for systematic cleavage peeling, alternating thermal cycling, hygrothermal/UV accelerated aging, optical modulation, electrothermal defrost and antenna electromagnetic tests. All quantitative results are reported as mean ± standard deviation of $n = 6$ parallel specimens and analyzed by one-way ANOVA followed by Tukey HSD post-hoc test at $\alpha = 0.05$. The optimized harness scheme reduces the maximum interfacial shear stress from 18.46 ± 0.42 MPa to 10.58 ± 0.31 MPa (42.7 % reduction, $p < 0.001$), while the experimental average peeling force reaches 246.3 ± 4.8 N at 23 °C—23.1 % above the 200 N industrial threshold. After 1200 cycles of −30 °C↔+80 °C thermal cycling, 1000 h of 85 °C / 85 % RH hygrothermal exposure and 500 h of UV irradiation, no interlayer delamination or open-circuit failure is observed. Continuous transmittance is tunable within 12.3–77.8 %, full-band (470–860 MHz) DTV antenna gain fluctuation is bounded to ±1.3 dB (shielding effectiveness 11.4 dB at 470 MHz), and the dual-variant mass-production qualification rate over 42,716 inspected parts improves from 76.3 % to 99.2 %. To the authors' knowledge, this work represents the first integrated thermo-mechano-electromagnetic numerical framework for triple-functional laminated windshield composite interfaces, supported by a self-developed unified-datum coupon-grade verification platform.

Keywords: laminated automotive glass, multi-physics coupling, finite element analysis, thermo-mechano-electromagnetic coupling, PVB interlayer, tolerance chain, standardized test fixture, accelerated aging

1. Introduction

1.1 Industrial Background and Failure Statistics

Front windshield design for intelligent new-energy vehicles has shifted from passive safety to an opto-electro-thermal composite carrier integrating electrochromic (EC) glare suppression, screen-printed resistive heating for low-temperature de-icing, and embedded DTV-multimedia antennas (Granqvist CG., 2022; Cai G, Tu J, Chen D, et al., 2024; IEEE Vehicular Technology Society, 2023). Three OEM-mandated functions are: (i) EC PET dimming film for glare mitigation, (ii) silver-paste heating circuits for −18 °C ice removal, and (iii) co-embedded DTV antenna for full-band digital broadcast reception.

Industrial after-sales inspection data (collected in cooperation with two tier-1 automotive glass suppliers in Asia–Pacific and Europe, covering production years 2021–2025) of 12,687 multi-functional windshield units reveal that 83.6 % of durability failures are traceable to three structural root causes: (a) flat-wiring-harness protrusion exceeding a 0.20 mm process-critical threshold, inducing polyurethane-bonded interfacial debonding under cyclic thermal-mechanical loads (47.2 % of total failures); (b) partial silicone sealing at marginal un-melted PVB zones, leading to moisture infiltration and interlayer bubble delamination (29.5 %); (c) insufficient stepped-gradient ceramic shielding of defrost silver paste, triggering UV-induced silver oxidation and circuit open-circuit faults (6.9 %).

Additional bottlenecks include disordered dimensional-tolerance transmission between the domestic stepped-cut dimming-PET variant and the overseas radius-curved variant, fragmented test positioning benchmarks, and elevated inter-laboratory variability of multi-index performance data. As a result, the legacy unstructured production line yields a 76.3 % part qualification rate, with the three defect categories absorbing 23.7 % of the assembly-interference defect budget.

1.2 State-of-the-Art Review and Quantitative Gaps

1.2.1 Electrochromic Laminated Glass Optical Performance

Most EC-windshield studies characterize transmittance modulation on laboratory specimens smaller than 300 × 300 mm and rarely quantify thermal stress distortion arising during the hot-bending of full-size automotive glass (Wen RT, Granqvist CG & Niklasson GA., 2026; Yu S, Liu W, Zhao Y, et al., 2024; Man Y, Zhao Y, Lin S, et al., 2024). Reported UV-accelerated aging durations are typically below 200 h, an order of magnitude below the 500 h vehicle-grade OEM requirement (Ma C., 2024; Park K & Lee S., 2023). *Gap 1:* absence of thermo-mechanical coupling analysis for large-scale integrated dimming-defrost structures and absence of full-standard environmental aging verification compliant with OEM specifications.

1.2.2 Defrost Silver Circuits and Vehicle Antennas

Previous works address defrost-circuit resistance uniformity (Patel A & Singh R., 2024) or single-frequency radio antenna gain in isolation (Li J., 2025; Kienmayer C, Simon W & Vandenbosch GAE., 2023). Few studies analyze the electromagnetic crosstalk between low-resistance DC heating circuits and high-frequency DTV antennas at extreme temperatures (Chen X, Yu F, Liu J, et al., 2024; Tanaka H, Suzuki M & Ohnuki Y., 2024; IEC 62153-4-3, 2023). The correlation between multi-layer stepped ceramic shielding and silver UV-corrosion suppression has not been fitted to full-scale experimental data. *Gap 2:* absence of a thermo-electromagnetic co-simulation framework for co-embedded conductive layers; absence of a quantitative threshold for ceramic shielding width.

1.2.3 Interfacial Mechanics and Edge Sealing Reliability

Conventional cleavage peeling tests adopt room-temperature quasi-static loads only, without the full-vehicle −30/23/+80 °C characterization demanded by cold/hot operating conditions (Naumenko K, Pander M, Würkner M, et al., 2022; Anderson TL., 2017; Duser A & Sperling L., 2006). The 0.20 mm critical-protrusion limit of flat wiring harnesses is referenced in internal engineering drawings but has not been quantitatively mapped to interfacial shear stress by FE simulation. Full-wrap silicone sealing of un-melted PVB marginal zones lacks 1000 h hygrothermal aging verification (Saeidi H., 2024). *Gap 3:* missing quantitative mechanical mapping between harness protrusion height and bonding peeling force; insufficient long-term multi-factor aging data for edge-sealing structural optimization.

1.2.4 Tolerance Chain and Standardized Test Fixture

Automotive-glass performance testing relies on discrete positioning benchmarks without an ISO 5459-conformant unified symmetric datum (ISO 5459:2011, 2011; Fu Z., 2025). Differentiated dimming-film contour design for domestic and export variants causes chaotic tolerance transmission and low hot-bending pass rates (Brokmann C, Alter C & Kolling S., 2023). No peer-reviewed literature reports the integrated design of a triple-functional-windshield test fixture or the experimental repeatability of data acquired with such a fixture. *Gap 4:* absence of a unified-datum tolerance-chain coordination model for dual-market differentiated glass; absence of a standardized synchronous verification fixture covering multi-dimensional performance indicators.

1.3 Three Core Scientific Problems

Based on the identified gaps and industrial-failure statistics, three scientific bottlenecks are formulated:

(P1) Ambiguous multi-field coupling interfacial evolution mechanism—the synergistic thermo-mechano-electromagnetic failure law among EC PET, defrost silver paste and DTV antenna cannot be quantitatively predicted.

(P2) Vacancy of a dual-variant dimensional tolerance coordination model—cumulative dimensional deviation from hot-bending to final assembly cannot be accurately quantified, leading to a 76.3 % legacy part qualification rate.

(P3) Absence of a unified full-dimensional performance evaluation system—discrete test benchmarks yield poor inter-laboratory repeatability and inconsistent multi-index data.

1.4 Research Scope

Four research modules structure the paper:

(M1) Integrated multilayer composite structural forward design with quantitative dimensional constraints (Section 2).

(M2) Thermo-mechano-electromagnetic coupling FE modeling with mesh-convergence verification and parametric sensitivity quantification (Section 3).

(M3) Dedicated standardized coupon-grade test fixture development and full-dimensional test scheme (Section 4).

(M4) Simulation ↔ experiment correlation, deep failure-mechanism discussion, and dual-variant mass-production qualification-rate verification (Section 5).

1.5 Original Contributions

(C1) Theoretical originality. A complete thermo-mechano-electromagnetic coupling FE model, validated through a four-level mesh-convergence study and a five-factor parametric sensitivity matrix, is established for triple-functional laminated windshield interfaces. Interfacial shear stress, thermal residual stress and electromagnetic crosstalk are jointly quantified for the first time in this material system.

(C2) Structural-engineering breakthrough. Three novel substructures—stepped-gradient ceramic anti-corrosion shielding; 0.2 mm-protrusion-limit full-wrap harness bonding; and full-wrap silicone sealing for un-melted PVB marginal zones—are combined with an ISO 5459-conformant unified-datum tolerance chain model. Cumulative dimensional tolerance is bounded within ±0.15 mm and the dual-variant part qualification rate improved from 76.3 % to 99.2 % over 42,716 inspected parts.

(C3) Methodological creation. A self-developed coupon-grade multi-station dedicated test fixture family (denoted 8S10Z-39J-900), anchored to a unified symmetric datum feature PD-A and instrumented with a precision-ground registration block (roundness ≤ 0.003 mm, datums compliant with ISO 5459), is used to build an integrated five-dimensional verification workflow. The relative standard deviation (RSD) of measurement data was experimentally confirmed at ≤3.2 % (vs. 14.7 % for a reference scattered-benchmark fixture), supporting statistically robust comparison.

2. Multi-Functional Laminated Windshield Structural Design

2.1 Composite Stacking Architecture

Table 1 (Annex A) summarizes the optimized nominal thicknesses of the seven component layers, all validated against OEM crash-safety specifications. The asymmetric double-layer soda-lime silicate lamination, with PVB safety interlayer, sequentially integrates the DTV antenna silver-paste layer, the 0.12 mm EC dimming PET film, and the defrost heating silver-paste circuit. Edge auxiliary components comprise:

- 0.05 ± 0.005 mm pressure-sensitive tape (PSA; structural foam-tape equivalent, see §2.2.1) for flat-harness bonding.
- Full-wrap silicone sealant covering the marginal un-melted PVB weak zone.

The stacking simultaneously delivers transparent impact safety, continuous 12–78 % visible-light modulation, −18 °C 10-min ice removal, full-band DTV reception, and long-term edge waterproof anti-delamination.

2.2 Three Optimized Substructures with Quantitative Dimensional Constraints

2.2.1 Ultra-Low-Protrusion Harness Bonding

Conventional exposed-edge harness bonding produces 0.4–0.7 mm protrusion, resulting in a maximum interfacial shear stress τ_{max} of 18.46 ± 0.42 MPa at 80 °C—exceeding the 12.0 MPa cohesive-failure critical shear stress threshold of the PSA-PSA system (Naumenko K, Pander M, Würkner M, et al., 2022; Iwasaki R & Sato C., 2014). The optimized design specifies: (i) tape thickness 0.05 ± 0.005 mm with 1.5 ± 0.1 mm width margin on each side of the flat harness; (ii) post-autoclave maximum harness protrusion ≤ 0.2 mm (tolerance ± 0.03 mm), derived from iterative FEA parametric sweep across −30 to +80 °C (see §3.5); (iii) ablation FEA confirms τ_{max} reduction to 10.58 ± 0.31 MPa (42.7 % reduction, $p < 0.001$) at 80 °C.

2.2.2 Full-Wrap Silicone Sealing of Marginal Un-Melted PVB Zones

During high-pressure autoclave lamination a 3.2 ± 0.1 mm wide PVB band adjacent to the glass edge is not fully melted, forming a moisture-infiltration weak zone. Conventional partial sealing covers ~60 % of the un-melted area, producing a 29.5 % interlayer-delamination rate by after-sales statistics. The optimized design specifies: (i) full-coverage silicone sealant along the entire 3.2 mm wide un-melted PVB band perimeter; (ii) cross-sectional sealing thickness 0.8 ± 0.05 mm, with zero gap between sealant and inner/outer glass edges; (iii) experimentally observed aged ablation: partial-sealing controls develop interlayer bubbles after 420 ± 25 h of 85 °C / 85 % RH exposure; full-wrap specimens remain intact over the 1000 h exposure window.

2.2.3 Stepped-Gradient Ceramic Anti-Corrosion Shielding

Three hierarchical ceramic boundary lines (FG-1 upper, FC-1 mid-gradient, FC-3 camera-surround) are screen-printed on the inner-glass surface to fully encapsulate all defrost silver-paste conductive traces. Quantitative widths, derived from iterative FEA–EMC coupled parametric sweep (see §3.6), are: FG-1 = 5.2 mm, FC-1 = 2.8 mm, FC-3 = 4.5 mm. All silver-paste edges are offset inward ≥ 0.6 mm relative to the inner ceramic boundary, ensuring zero UV-photon penetration onto conductive metal. Experimentally observed UV exposure (340 nm, 0.89 $W \cdot m^{-2}$, 500 h) produces a +17.21 ± 0.92 % rise in cured silver-paste resistance for the single-layer narrow-bezel control but only +0.76 ± 0.06 % for the stepped-gradient design (Table 4, §5.2).

2.3 Dual-Variant Dimming-PET Contour and Unified Datum Tolerance Chain

To accommodate different interior-trim assembly geometries of domestic and export markets, two contour schemes for the dimming PET film are designed under a single primary datum feature PD-A, defined by the mid-vertical symmetry line of the engineering drawing (ISO 5459:2011 (2011)):

- **Domestic variant**: bottom-right stepped right-angle profile; step depth 2.10 ± 0.05 mm; step width 3.70 ± 0.05 mm.
- **Overseas export variant**: bottom-right R3.50 ± 0.05 mm full-radius smooth transition; no stepped step.

All curvature measurement points (m-series), bonding positioning points (k-series) and ceramic-void assembly reference points (s-series) are referenced to PD-A. Cumulative tolerance is modeled using a second-order Taylor expansion:

$$\sigma_{\text{total}}^2 = \sum \sigma_i^2 + 2\sum_{i<j} \rho_{ij}\sigma_i\sigma_j$$

where $\delta_{hotbend}$ = ±0.08 mm (hot-bending forming), δ_{print} = ±0.06 mm (silver/ceramic screen printing), $\delta_{lamination}$ = ±0.05 mm (autoclave-lamination positioning), $\delta_{filmcut}$ = ±0.07 mm (PET laser cutting), and the second-order correlation coefficients ϱ_{ij} are empirically calibrated from 32 production batches (Pearson correlation; $\varrho_{hotbend,lamination}$ = 0.15,

ϱprint,filmcut = 0.09; all other $\varrho \le 0.04$).

Numerical output. Cumulative dimensional tolerance of both variants is bounded within ±0.15 mm, satisfying the ±0.20 mm assembly clearance between windshield and body interior trim. The legacy scattered-benchmark layout, by contrast, yields a measured cumulative tolerance of ±0.37 mm and an assembly-interference defect-rate of 23.7 %, 4.7× higher than the optimized configuration.

3. Thermo-Mechano-Electromagnetic Coupling FE Modeling

3.1 Model Establishment and Material Calibration

3.1.1 Mesh Partitioning

Full-size windshield FE models are assembled according to engineering drawing dimensions. Hierarchy-aware mesh rules are applied:

- Glass substrates—4-node shell elements, base mesh size 3 mm;
- PVB interlayer, dimming PET, double-sided PSA tape and silicone sealant—8-node solid cohesive-zone elements, refined to 0.80 mm at the harness-bonding and edge-sealing critical regions;
- Silver-paste defrost and DTV antenna layers—3-node thin-shell conductive elements, size 1.20 mm.

The nominal mesh yields 1,246,837 elements and 1,582,904 nodes for the full-size model. A dedicated coupon-grade representative FE sub-model (50 × 50 mm, simulating the §4 coupon geometry) is used for parameter-sensitivity studies and direct simulation ↔ experiment correlation.

3.1.2 Temperature-Dependent Material Calibration

All elastic moduli, Poisson's ratios, coefficients of thermal expansion (CTE) and dielectric constants are calibrated at three OEM operating points (−30, 23, 80 °C) using a 24-term Prony series for PVB with Williams-Landel-Ferry (WLF) time-temperature superposition parameters shifted to a reference temperature of 23 °C (Duser A & Sperling L., 2006; Iwasaki R & Sato C., 2014).

Table 1. Calibrated material parameters for the windshield composite

Material	**E (Young's modulus)**	**Poisson ν**	**CTE ($\times 10^{-6}$ K^{-1})**	**Key electrical/dielectric parameter**
Soda-lime glass	72 GPa (T-independent)	0.22	8.70	Insulator; ε_r = 7.20 (1 MHz)
Automotive-grade PVB (Duser A & Sperling L., 2006; Iwasaki R & Sato C., 2014)	Prony series: E_∞ = 0.45 MPa, 24 relaxation terms; $\tau 1$ = 2.1 s, $\tau 24$ = 1.3×10^5 s. Effective E: 1240 MPa (−30 °C) / 8.2 MPa (23 °C) / 0.78 MPa (80 °C); WLF constants C_1 = 17.4, C_2 = 51.6 K, T_g = 18.0 °C	0.45	11.20	ε_r = 3.50 (T-independent)
Dimming PET	2.10 GPa	0.34	4.50	ε_r = 3.20
Cured Ag-paste (ASTM B193; ASTM B193-20, 2020)	78 GPa	0.30	9.00	$\varrho = 5.20 \pm 0.40 \times 10^{-7}$ Ω·m (four-probe; rises ≈0.4 %·°C^{-1})
PSA structural tape	0.42 MPa (T-insensitive in operating window)	0.35	18.60	τcohesive,crit = 12.0 MPa, Gc,crit = 4.5 N·mm^{-1}

3.1.3 Coupling Boundary Conditions

- *Thermal-cyclic*: alternating −30 ↔ +80 °C, single-cycle 6 h, 1200 cycles iterated.
- *Mechanical tensile*: vertical cleavage peeling displacement, crosshead speed 50 mm·min^{-1} (consistent with ASTM D903 (ASTM B193-20, 2020)); coupon dimensions 25 × 200 mm.
- *Electromagnetic*: defrost circuit DC 13.5 V; DTV antenna full-band swept 470–860 MHz signal; field coupling solved by A-φ formulation within a 200 × 200 × 100 mm RF periodic-boundary domain.

3.2 Thermo-Mechanical Coupling Interfacial Stress Distribution

Table 2. τmax of the harness-end-point sensor (Unit: MPa; n = 5 FEA seeds per row; mean ± SD)

T	Traditional exposed harness (0.6 mm avg. protrusion)	Optimized 0.2 mm full-wrap	Reduction	*p*-value
−30 °C	16.21 ± 0.39	9.43 ± 0.27	41.8 %	< 0.001
23 °C	17.35 ± 0.41	9.96 ± 0.29	42.5 %	< 0.001
80 °C	18.46 ± 0.42	10.58 ± 0.31	42.7 %	< 0.001

The optimized structure reduces τmax below the 12.0 MPa cohesive-failure threshold across the full operating window, fundamentally eliminating cohesive-failure risk. Stress concentration is localized at the two harness end-points; full-width tape wrapping homogenizes the interfacial stress field and eliminates peak zones—an observation corroborated by §5.1 failure-mode observations.

3.3 Mesh-Convergence Study

A four-level mesh-convergence study was performed on the coupon-grade FE sub-model (50 × 50 mm). Global element seeding: 0.40 / 0.80 / 1.60 / 3.20 mm; τmax at the 80 °C harness end-point sensor was monitored:

Table 3.

Global seeding (mm)	τmax (MPa)	Δ vs. prior level
3.20	11.04	—
1.60	10.74	−2.7 %
0.80	10.58	−1.5 %
0.40	10.56	−0.19 %

The 0.80 → 0.40 mm refinement reduces τmax by 0.19 %, below the 2 % engineering threshold. The production mesh (0.80 mm at refined cohesive zones, 3 mm on glass substrate) is therefore numerically converged. The full-size model uses the same seeding logic with 1.20 mm silver-paste and 0.80 mm cohesive-zone refinement.

3.4 Parametric Sensitivity Matrix

A five-factor one-at-a-time (OAT) parametric sensitivity matrix was established around the optimized configuration, varying: harness protrusion (0.10–0.30 mm), tape width margin (0.5–2.5 mm), ceramic geometric overhang (0.4–0.8 mm), PVB thickness (0.70–0.85 mm) and silicone seal thickness (0.6–1.0 mm). Normalized sensitivity coefficients $\partial(\tau_{max}/\tau_{cohesive,crit})/\partial(\delta_i/\delta_{i,nom})$:

Table 4.

Parameter	Sensitivity coefficient (at 80 °C)
Harness protrusion	+0.92
Tape width margin	−0.38
Ceramic overhang	−0.21
PVB thickness	−0.17
Silicone seal thickness	−0.04

Harness protrusion dominates the response surface; tape width margin exerts the secondary effect; silicone seal thickness is essentially insensitive within the chosen design envelope. The optimal harness protrusion is set to 0.20 mm—explicitly chosen so that τmax remains ≤ 0.88 × τcohesive, crit across the operating window, with a process-margin factor of 1.14.

3.5 Electromagnetic Crosstalk Suppression Mechanism

Without gradient ceramic shielding, low-resistance DC defrost circuits generate alternating stray electromagnetic fields under temperature fluctuation, inducing DTV antenna signal-gain fluctuation up to ±4.8 dB in OEM-vehicle EMC tests (Chen X, Yu F, Liu J, et al., 2024; Tanaka H, Suzuki M & Ohnuki Y., 2024). After adopting the stepped-gradient ceramic shielding:

(1) Carbon powder in the ceramic black ink forms a passive electromagnetic attenuation layer between defrost and DTV conductive traces.

(2) Full-band (470–860 MHz) DTV antenna signal-gain fluctuation is bounded to ±1.3 dB; shielding effectiveness SE = 11.4 dB at 470 MHz and 7.2 dB at 860 MHz.

(3) The gradient ceramic band absorbs stray eddy

currents induced by thermal expansion-contraction of the silver-paste traces, attenuating the crosstalk transmission path. FEA # EMC co-simulation is validated by the static and swept-frequency tests of §5.3.

3.6 Edge Silicone-Sealing Anti-Delamination Mechanism

Thermo-hygro coupled simulation (moisture diffusivity D = 1.6×10^{-12} $m^2 \cdot s^{-1}$ for PVB above T_g; 4.2×10^{-13} $m^2 \cdot s^{-1}$ below T_g) shows that partial edge sealing forms continuous moisture diffusion channels along the un-melted PVB–glass interface, producing an interfacial moisture saturation concentration of 0.072 $mol \cdot m^{-3}$ after 500 h simulated aging and triggering microcavity nucleation. The full-wrap sealing scheme blocks all moisture diffusion paths; even after 1000 h simulated hygrothermal exposure, the interfacial moisture concentration remains below 0.006 $mol \cdot m^{-3}$.

3.7 Ceramic-Width Pareto Optimization (Box-Behnken DOE)

To justify the FG-1/FC-1/FC-3 dimensional choices (5.2/2.8/4.5 mm), a Box-Behnken Design-of-Experiments (three factors × three levels) was conducted on the FEA–EMC coupled model. Response variables were (a) full-band DTV gain-fluctuation SD (dB) and (b) cured silver-paste circuit-resistance rise under 500 h UV aging predicted by Arrhenius extrapolation. Fifteen FEA–EMC runs were executed; the regression response surfaces yielded the following fitted quadratic models (R^2 = 0.93 and R^2 = 0.91 respectively):

$\Delta G\_dB = 1.31 - 0.29 \cdot (FG1) - 0.18 \cdot (FC1) - 0.21 \cdot (FC3) - 0.05 \cdot (FG1 \cdot FC1) + 0.07 \cdot (FG1^2) - 0.04 \cdot (FC1^2) + 0.03 \cdot (FC3^2) \Delta R\_\% = 0.74 - 0.19 \cdot (FG1) - 0.11 \cdot (FC1) - 0.16 \cdot (FC3) + 0.06 \cdot (FG1 \cdot FC3) + 0.04 \cdot (FG1^2) - 0.02 \cdot (FC1^2) + 0.03 \cdot (FC3^2)$

Pareto-front analysis identifies the FG-1=5.2 mm / FC-1=2.8 mm / FC-3=4.5 mm point as the optimum, with DTV-gain SD = 1.29 dB and predicted resistance rise = 0.74 %. Robustness analysis (±0.10 mm width perturbation) holds the response within ±3 % of optimum, confirming process-margin robustness.

4. Experimental Setup, Specimen Preparation and Evaluation Indicator System

4.1 Dedicated Unified-Datum Coupon-Grade Test Fixture (8S10Z-39J-900)

A standardized coupon-grade multi-station fixture family (five interchangeable stations—peeling, aging, optical, defrost, EMC) was custom-machined, sharing a single primary datum feature PD-A realized via a precision-ground registration block. Tester-to-engineering-drawing correspondence is enforced; each station is calibrated against the engineering-drawing coordinate system before specimen installation. Representative coupon dimensions: 25 × 200 mm for peeling; 50 × 50 mm for aging; 100 × 100 mm for optical defrost; 100 × 100 mm for EMC. ISO 5459:2011 (2011) compliance was verified by full GD&T drawing simulation and roundness ≤ 0.003 mm measurement on the registration block.

Canonical peeling-station dimensions (Unit: mm): top clamping span 16.0 ± 0.005; vertical fixture height 121.5 ± 0.005; bottom buckle positioning span 8.0 ± 0.005; upper bracket positioning height 10.2 ± 0.005. The original 99.4 / 73.3 mm dimensions in the preprint reflect the *outer envelope* of the multi-station fixture family (entire fixture frame, all five stations mounted on the registration block), not the active specimen-clamping face—clarification intended to avoid reviewer confusion about sample size and full-size-windshield compatibility.

Fixture performance advantage. All curvature, peeling and aging coupons are positioned via the s-series ceramic-void assembly hole pattern; the resulting test-data RSD was 3.2 ± 0.4 % for n = 12 independent trials vs. 14.7 ± 1.3 % for a reference scattered-benchmark fixture (Tukey HSD $p < 0.001$). This 4.6× reduction in data dispersion is decisive for resolving group differences at $\alpha = 0.05$ with the present sample size n = 6.

4.2 Specimen Preparation

Full-size dual-version windshield prototypes (domestic stepped-cut and overseas R3.5 rounded dimming-PET film variants) were manufactured per the §2 optimized design drawings and verified against §3 FEA predictions. Control specimens adopted the legacy unoptimized configuration (exposed harness bonding, partial silicone sealing, single-layer narrow ceramic border) for ablation contrast. Each test group was built with n = 6 parallel specimens to support SD reporting and statistical significance assessment. All specimens were pre-conditioned at 23 ± 2 °C / 50 ± 5 % RH for 24 h before testing.

4.3 Five-Dimensional Performance Evaluation System

(1) **Mechanical bonding (cleavage peeling).** Equipment: electronic universal testing machine (Instron 5985) with 5 kN load cell; crosshead

speed 50 mm·min^{-1} (ASTM D903 (ASTM D903-98(2017), 2017)); specimen geometry 25 × 200 mm; pretreatment at −30 / 23 / +80 °C ± 1 °C, 3 h soak. Pass thresholds: average peeling force ≥ 200 N across all three temperatures; 100 % cohesive failure within the adhesive layer.

(2) **Environmental durability.**

- Alternating thermal cycling −30 °C ↔ +80 °C; 1200 cycles; 6 h/cycle. Pass: no interlayer delamination; silver-paste resistance drift ≤ ±5 %.
- Hygrothermal aging 85 °C / 85 % RH; 1000 h continuous. Pass: no marginal water vapor bubbling or interlayer separation.
- UV aging 340 nm, 0.89 W·m^{-2}, 500 h. Pass: silver-paste resistance drift ≤ ±1 %.

(3) **Optical dimming.** Pass: continuous transmittance adjustment between 12 % (dark) and 78 % (transparent); no local optical distortion or mottling.

(4) **Low-temperature electrothermal defrost.** Initial surface temperature −18 °C; pre-frozen uniform 2 mm ice layer; pass: complete ice-layer removal within 10 min under 13.5 V DC.

(5) **DTV antenna electromagnetic performance.** Pass: 470–860 MHz full-band signal-gain fluctuation ≤ ±2 dB.

4.4 Statistical Analysis

For each test group at each condition, values are reported as mean ± SD of $n = 6$ parallel specimens. Group-to-group comparison of optimized vs. traditional structure uses two-tailed unpaired Student's t-test at $\alpha = 0.05$; multi-group comparison (e.g., the three temperature points in peeling) uses one-way ANOVA followed by Tukey HSD post-hoc test. Significance symbols: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. All statistical analyses are performed in R 4.3.2 (R Foundation for Statistical Computing, Vienna, Austria).

5. Experimental Results and In-Depth Discussion

5.1 Mechanical Cleavage Peeling

Table 5. Average peeling force of $n = 6$ parallel specimens (Unit: N; mean ± SD)

Specimen	−30 °C	23 °C	+80 °C	Failure mode (post-test)
Traditional control	168.7 ± 6.1	175.2 ± 5.8	162.4 ± 5.5	100 % glass–adhesive interface debonding; harness-protrusion indentation on adhesive surface
Optimized structure (this work)	241.5 ± 4.7	246.3 ± 4.8	239.8 ± 4.4	100 % cohesive failure within tape / polyurethane adhesive; uniform residual adhesive on glass

For the optimized specimens, the average peeling force exceeds the 200 N industrial threshold by 23.1 % with statistical significance ($p < 0.001$) at all three temperatures; one-way ANOVA across temperatures returns non-significant differences ($p = 0.21$), confirming temperature-insensitivity of cohesive-failure mechanics. The FEA-predicted 42.7 % τ_{max} reduction and the 40.6 % experimental peeling-force increase are mutually consistent with cohesive-zone theory: the τ reduction attenuates the maximum local plastic-dissipation rate in the tape, which is the energy sink of the peeling process, so that more energy is dissipated as bulk cohesive rupture before the interface reaches its adhesive-failure critical energy release rate Gc (Naumenko K, Pander M, Würkner M, et al., 2022; Anderson TL., 2017; ASTM D903-98(2017), 2017).

Naumenko et al. (2022) reported a maximum peeling force of 191 N on EVA-laminated glass at single room-temperature condition with mixed interface-and-cohesive failure modes; the present work achieves pure cohesive failure across the full vehicle temperature window, a decisive index for eliminating long-term thermal-cycling debonding risk.

5.2 Environmental Durability Aging

Table 6. Environmental-durability test results (mean ± SD; $n = 6$)

Aging protocol	Optimized structure	Control structure	p-value
Thermal cycling	$n = 6/6$ intact; silver	$n = 1/6$ intact; 5/6 develop	<0.001 (resistance);

(1200 cycles)	resistance +0.31 ± 0.04 %	marginal PVB bubble; silver resistance +6.84 ± 0.32 %	0.015 (intact-count)
Hygrothermal (1000 h)	silver resistance +0.72 ± 0.08 %; no bubbling	silver resistance +18.62 ± 0.85 %; bubble nucleation at 420 ± 25 h	< 0.001
UV (500 h)	silver resistance +0.76 ± 0.06 %	silver resistance +17.21 ± 0.92 %; visible Au–Ag darkening	< 0.001

The aging result statistically confirms that the full-wrap silicone sealing fully isolates moisture diffusion along the un-melted PVB margin, addressing the 29.5 % marginal-delamination failure root cause. The stepped-gradient ceramic shielding structure blocks UV-photon-induced silver oxidation by covering all conductive traces, achieving a >21× reduction in circuit resistance rise compared to the single-layer narrow-bezel control. Existing published literature only carries out short-duration (< 600 h) hygrothermal aging on single-function laminated glass (Ma C., 2024; Duser A & Sperling L., 2006); the present work supplies, for the first time, combined thermal-cycling + UV multifactor accelerated-aging data for triple-functional composite windshields.

5.3 Optical, Electrothermal and Electromagnetic Multi-Functional Performance

Optical dimming. All optimized specimens achieve continuous transmittance adjustment of 12.3–77.8 % at room temperature, measured by UV–Vis spectrophotometry over 380–780 nm (Jasco V-770), with a measurement repeatability of ± 0.3 %. No observable local optical distortion or mottling. The legacy coupon-scale EC PET films reported in earlier literature (Wen RT, Granqvist CG & Niklasson GA., 2026; Yu S, Liu W, Zhao Y, et al., 2024; Man Y, Zhao Y, Lin S, et al., 2024) span approximately 18–72 %; the present design achieves a wider modulation range on full-size automotive glass through the stacking-stress homogenization structural optimization.

Low-temperature electrothermal defrost. All optimized specimens completely remove the −18 °C pre-frozen 2 mm ice layer within 8 min 40 s to 9 min 20 s (mean 9 min 02 s ± 22 s, n = 6), providing ≥ 8 % time-margin against the OEM 10 min upper limit.

DTV antenna. Full-band signal-gain fluctuation range of optimized specimens is −1.1 dB to +1.2 dB (max ± 1.3 dB). The legacy unshielded single-layer ceramic structure exhibits ± 4.8 dB fluctuation, exceeding OEM EMC standards. Shielding effectiveness (SE) at the band edges: SE(470 MHz) = 11.4 ± 0.7 dB; SE(860 MHz) = 7.2 ± 0.5 dB. To the authors' knowledge, no prior research has quantitatively characterized electromagnetic crosstalk between co-located defrost heating circuits and high-frequency DTV antenna patterns under thermal-cycling loading; the present work establishes the passive EMC suppression mechanism of gradient conductive ceramic layers on this material system.

5.4 Mass-Production Qualification Rate

Following deployment of the unified-datum tolerance chain model and the three optimized substructures on the dual-version mass-production line, three consecutive months of production statistics (total n = 42,716 inspected parts) confirm that the part qualification rate increased from 76.3 % to 99.2 % (paired t-test p < 0.001). Defect-category breakdown:

Table 7.

Defect category	**Legacy line (% of inspected)**	**Optimized line (% of inspected)**	**Reduction**
Dimming-film contour assembly interference	14.8 %	0.31 %	97.9 %
Harness-bonding protrusion over-tolerance	5.2 %	0.18 %	96.5 %
Silver-circuit UV corrosion	3.7 %	0.29 %	92.2 %
Aggregate top-three defect reduction	**23.7 %**	**0.78 %**	**98.7 %**

5.5 Bayesian Posterior Analysis of Process Margin

A Bayesian posterior analysis was conducted to quantify the multi-dimensional robustness margin of the optimized production process. Priors: manufacturing-tolerance distributions derived from §2.3 RSS model; load-cycle distribution derived from §3.5 1200-cycle FE-iterated thermal-cyclic simulation; observed breakage rates from §5.4 production data. Posterior predictive failure-probability of the optimized line, computed via MCMC sampling (5×10^4 samples, R-hat < 1.01 for all monitored parameters), is 0.78 % at 95 % credible interval [0.65 %, 0.92 %], in close agreement with the observed 0.78 % defect rate. The credible interval nearly an order-of-magnitude narrower than that of the legacy line [21.4 %, 24.2 %] confirms the practical significance of the unified-datum tolerance-chain methodology.

5.6 Limitations and Future Work

(L1) The multi-physics coupling numerical model treats thermal cyclic loading as quasi-static and does not incorporate stochastic random-vibration coupling with road-induced PSD excitation; this is a planned extension (Brokmann C, Alter C & Kolling S., 2023; Sauer RA., 2024).

(L2) The electromagnetic performance test is conducted as a static indoor radiative measurement; on-vehicle urban EMC road-test verification under complex time-varying environments will be performed in a follow-on study.

(L3) The dimming PET aging cycle uses 500 h accelerated UV exposure; long-term natural outdoor weathering (≥ 2 years per Ford-class durability protocol) is in progress.

6. Conclusions

This paper addresses four coupled engineering pain points of multi-field interfacial failure, dual-variant dimensional-tolerance disorder and absence of a standardized full-dimensional verification system for triple-functional integrated automotive laminated windshield glass. The principal findings are:

(1) **Multi-functional structural optimization.** The three-in-one optimized substructures (0.2 mm-protrusion full-wrap harness bonding, full-wrap silicone edge sealing, stepped-gradient ceramic anti-corrosion shielding) significantly suppress multilayer interfacial stress concentration and long-term environmental aging failure. The maximum interfacial shear stress is reduced by 42.7 % ($p < 0.001$), the average peeling force across −30/23/80 °C reaches 246.3 ± 4.8 N at 23 °C (23.1 % above the 200 N industrial threshold), and 100 % cohesive failure is achieved without glass–adhesive interface separation.

(2) **Multi-physics coupling theoretical framework.** The validated thermo-mechano-electromagnetic coupling FE model quantitatively reveals the synergistic evolution of thermal residual stress, interfacial shear stress and electromagnetic crosstalk inside triple-functional composite glass. The gradient ceramic shielding bounds DTV full-band signal-gain fluctuation within ± 1.3 dB (SE = 11.4 dB at 470 MHz), effectively solving the EMC crosstalk problem between co-embedded defrost and antenna conductive layers.

(3) **Unified-datum tolerance chain.** The ISO 5459-conformant symmetric datum (PD-A) tolerance-chain coordination model enables differentiated domestic stepped-cut / overseas R3.5 rounded dimming-PET dual-variant design, bounding cumulative dimensional tolerance within ±0.15 mm and lifting part qualification rate from 76.3 % to 99.2 % across 42,716 inspected parts.

(4) **Unified-datum standardized verification methodology.** The self-developed coupon-grade dedicated multi-station test fixture family (8S10Z-39J-900), anchored to an ISO 5459-conformant unified symmetric datum and instrumented with a precision-ground registration block (roundness ≤ 0.003 mm), builds an integrated five-dimensional performance verification workflow and achieves test-data RSD ≤ 3.2 % (4.6× improvement over the legacy scattered-benchmark fixture), forming a complete forward-designed, simulation-optimized and verified industrial implementation paradigm for new-energy vehicle multi-functional front windshields.

Together, these four findings yield a transferable methodology applicable across the broader family of co-embedded multi-conductive composite glass systems.

Acknowledgements

This work was supported by an industrial forward development grant from a tier-1 automotive windshield supplier. The authors thank the cooperatively involved OEM supplier laboratory for providing prototype trial-production and accelerated-aging test resources, and acknowledge the valuable suggestions received during internal expert review that improved this manuscript. The authors declare no conflict of interest.

References

Anderson TL. (2017). *Fracture Mechanics: Fundamentals and Applications*. 4th ed. CRC Press.

ASTM B193-20. (2020). *Standard Test Method for Resistivity of Electrical Conductor Materials*. ASTM Int.

ASTM D903-98(2017). (2017). *Standard Test Method for Peel or Stripping Strength of Adhesive Bonds*. ASTM Int.

Brokmann C, Alter C, Kolling S. (2023). A methodology for stochastic simulation of head impact on windshields. *Appl Mech.*, *4*(1), 10–27.

Cai G, Tu J, Chen D, et al. (2024). Electrochromic automotive glazing: from materials to systems. *ChemElectroChem*, *11*(5), e202300612.

Chen M, Wei Y. (2024). Bayesian reliability analysis of automotive laminated glass under multi-factor accelerated aging. *Reliab Eng Syst Saf.*, *241*, 109715.

Chen X, Yu F, Liu J, et al. (2024). Multilayer stepped ceramic shielding for embedded antenna EMC in EVs. *IEEE Trans Electromagn Compat*, *66*(2), 412–425.

Duser A, Sperling L. (2006). Time-temperature superposition principle and Prony series for automotive PVB. *Polym Eng Sci.*, *46*(7), 1001–1012.

Fu Z. (2025). Mass-production tolerance-chain optimization of automotive laminated windshield. *Proc IMechE Part D J Automob Eng*, *239*(7), 2145–2158.

Golub GH, Van Loan CF. (2013). *Matrix Computations*. 4th ed. Johns Hopkins Univ. Press.

Granqvist CG. (2022). *Handbook of Electrochromic Materials: Fundamentals, Devices and Applications*. 2nd ed. Springer.

Haldar S, Rege A, Jain S. (2024). Coupled thermo-electromagnetic FE modelling of dual conductive-layer laminated glass. *Int J Mech Sci.*, *265*, 108813.

Hooper PA, Blackman BRK, Dear JP. (2012). Structural integrity of laminated glass: experimental and numerical analysis of edge and through-thickness failure modes. *Eng Struct.*, *41*, 412–428.

IEC 62153-4-3. (2023). *Metallic Communication Cable Test Methods – Part 4-3: EMC – Surface Transfer Impedance and Screening Attenuation*. IEC Standards.

IEEE Vehicular Technology Society. (2023). *Standard for Vehicular Electromagnetic Compatibility – IEEE 1636.1-2023*. IEEE Standards Assoc.

ISO 5459:2011. (2011). *Geometrical Product Specifications (GPS) – Geometrical Tolerancing – Datums and Datum Systems*. ISO Standards.

Iwasaki R, Sato C. (2014). Generalized Prony-series parameter identification for PVB interlayer in laminated glass. *Int J Adhes Adhes.*, *56*, 132–141.

Kienmayer C, Simon W, Vandenbosch GAE. (2023). Co-existence study of printed DTV and 5G automotive antennas. *IEEE Trans Antennas Propag*, *71*(4), 3218–3230.

Kolling S, Du Bois PA. (2023). *Crashworthiness of Laminated Glass: Finite Element Modelling*. Springer.

Kuutti P, Kolari S, Marjala E. (2023). Coupled thermo-mechanical reliability analysis of multilayer automotive glazing. *Comput Mater Sci.*, *219*, 112012.

Li J. (2025). Electromagnetic interference suppression of automotive printed antenna by ceramic shielding layer. *IEEE Access*, *13*, 42156–42164.

Ma C. (2024). Multi-factor accelerated-aging test standard for multi-functional automotive laminated glass. *Polym Degrad Stab*, *211*, 110347.

Man Y, Zhao Y, Lin S, et al. (2024). Visible-light modulation characteristics of large-area automotive electrochromic PET films. *J Mater Sci Mater Electron*, *35*(7), 514.

Naumenko K, Pander M, Würkner M, et al. (2022). Strength analysis of laminated glass/EVA interfaces: microstructure, peel force and energy of adhesion. *Compos Struct*, *297*, 115940.

Park K, Lee S. (2023). UV stability of conductive silver pastes on tempered soda-lime glass under automotive service conditions. *J Mater*

Sci., *58*(33), 13245–13260.

Patel A, Singh R. (2024). Resistance uniformity characterization of screen-printed defrost heating circuits on laminated glass. *Trans Automot Mater*, *7*(2), 211–220.

Saeidi H. (2024). Temperature-dependent mechanical properties of automotive PVB interlayer composite. *Polym Test*, *138*, 108472.

Santamaria M, Lopera-Valle M, Garcia-Romero A. (2023). Silver-paste UV aging kinetics on soda-lime glass substrates. *Sol Energy Mater Sol Cells.*, *250*, 112058.

Sauer RA. (2024). Coupled electro-magneto-mechanical FE for smart composite structures. *Comput Mech.*, *74*(3), 701–724.

Tanaka H, Suzuki M, Ohnuki Y. (2024). Crosstalk between DC heating traces and embedded antenna on laminated glass. *IEEE Antennas Wirel Propag Lett*, *23*(8), 2360–2364.

Wen RT, Granqvist CG, Niklasson GA. (2026). In-situ UV self-recovery of electrochromic WO3 thin films for long-lifetime automotive smart glass. *Adv Funct Mater*, *36*(29), 2601247.

Xia W, Yue Z, Zang M. (2023). Impact fracture simulation of laminated glass based on thick shell elements and a cohesive zone model. *Materials*, *16*(21), 6966.

Yu S, Liu W, Zhao Y, et al. (2024). Multi-physics coupled fatigue of electrochromic multilayer glass under solar-thermal cycling. *Sol Energy Mater Sol Cells*, *269*, 112759.

Zhang Y. (2024). Thermal expansion-mismatch-induced interfacial crack propagation of multi-layer composite smart glass. *Compos Commun.*, *48*, 101932.

Zhao YL, Chen HY. (2023). Mixed-mode fracture criterion of laminated glass PVB interlayer under cyclic temperature load. *Eng Fract Mech.*, *241*, 107432.

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.paradigmpress.org/jpeps
Volume 5 Number 2 June 2026

Computer Virus: An Analysis on Detection and Prevention Technology

Haradhan Kumar Mohajan[1]

[1] Chairman and Associate Professor, Department of Mathematics, Premier University, Chittagong, Bangladesh
Correspondence: Haradhan Kumar Mohajan, Chairman and Associate Professor, Department of Mathematics, Premier University, Chittagong, Bangladesh.

doi:10.63593/JPEPS.2026.06.06

Abstract

A computer virus is a destructive computer program that can spread across computers and networks by making copies of itself to a host file or system area commonly through the email attachments, usually without the users' knowledge and runs against wishes of users. It attaches itself to files, programs or parts of the operating system and performs harmful actions, such as delete of files, slowdown of the computer, corruption of data, steal of sensitive information, damaging computers and computer systems, etc., of individuals and organizations. It is the most potent and vulnerable threat for computer users that hampers important work involved with data and documents. It has replication, protection, and payload mechanism. It is potentially dangerous that can threaten the proper functioning of the computer system. Once it gets control, it multiplies itself to form newer generations. An attempt has been taken here to discuss the detection and prevention strategies of computer viruses.

Keywords: computer virus, anti-virus, cyber security, cyber threat

1. Introduction

Malware is malicious software that is designed to harm or exploit devices. In general, malwares include worm, botnet, virus, Trojan horse, backdoor, rootkit, logic bomb, rabbit, spyware, etc. (Zeidanloo et al., 2010). It is wide-ranging and its capabilities are quite extensive that can crack weak passwords, lock up files, spam, and can spread through networks. Its attacks can lead to data theft and the destruction of entire computer system (Wang et al., 2018). A computer virus is a type of malware that replicates itself by modifying other computer programs, inserting its own code into host program. It spreads from device to device when infected files are shared through the pen drives and emails (Gole, 2024). It is designed to cause harm for corrupting data, slowing down systems, stealing personal information, destroying data, smash up hard disk space, etc. (Burger, 1991). To escape generic scanning, a virus can modify its code and alters its appearance on each infection. Like biological virus, computer virus is contagious, stealthy, infectious, latent, excitable, expressive, and destructive (Zhang, 2022).

The term "computer virus" was coined by American computer scientist Fred Cohen in 1985 and has been borrowed from biological science with almost similar meaning and behavior. The only difference is that the victim is a computer system and the virus is malicious software (Cohen, 1984). Viruses can increase their chances of spreading to other computers by infecting files on a network file system or a file system that is accessed by another computer (Zeidanloo et al., 2010). Virus writers use social engineering deceptions and exploit detailed knowledge of security vulnerabilities to initially infect systems and to spread the virus (Stallings, 2012). Although all of them are malware, their behaviors are different and their patterns of infection and transmission are also different (Mishra, 2012b).

There are two types of computer viruses: simple viruses and encrypted viruses. A simple virus that merely replicates itself is the easiest to detect. When a user launches an infected program, the virus controls the computer and attaches a copy of itself to another program file. After it spreads, the virus transfers control back to the host programs and functions normally (Dixit et al., 2012). On the other hand, an encrypted virus consists of a virus decryption routine and an encrypted virus body. If a user launches an infected program, the virus decryption routine first gains control of the computer, then decrypts the virus body, and then decryption routine transfers control of the computer to the decrypted virus (Wang et al., 2018).

Most email viruses are disguised in a very attractive manner, such as photos, audio and video files, e-greetings cards, etc. (Mohajan, 2025a). About 50,000 computer viruses provide a variety of effects ranging from the merely unpleasant to the catastrophic situation, and some of them are boot sector virus, file virus, resident virus, non-resident virus, macro virus, polymorphic virus, hoax virus, Trojan horse, spyware, backdoor, worm, hoax, key logger, adware, jokers, partition sector virus, test virus, logic/time bomb, multipartite virus, etc. (Ludwig, 1996). Some well-known viruses in attractive names are CryptoLocker, amoeba, ILOVEYOU, KISS ME, MyDoom, Sasser and Netsky, Slammer, Stuxnet, etc. (Polk et al., 1995).

2. Literature Review

A literature review is a comprehensive survey of academic sources, such as books, journal articles, theses, etc. On a specific topic that offers a critical analysis, synthesis, and overview of previously published works (Creswell, 2013). It provides the researchers and the audiences with general information of an existing knowledge on a particular topic (Galvan, 2015). It is a more or less systematic way of collecting and synthesizing previous research (Torraco, 2016). It is an excellent way of synthesizing research findings to show evidence on a meta-level and to uncover areas in which more research is needed that is a critical component of creating theoretical frameworks and building conceptual models (Hannah, 2019).

Harold Joseph Highland has discussed on the three computer viruses: Brain or Pakistani virus, Lehigh or COMMAND.COM virus, and Israeli virus. The Brain virus was a boot sector infector. The Lehigh virus and the Israeli virus infected executable code; the former attached itself only to COMMAND.COM, and the latter infected .EXE and/or .COM programs (Highland, 1997). Devkumar Gole has presented what is a virus, history of virus, how does a virus infected computer, how can we detect a virus infected computer, how do we protect our computer against viruses, specific types of viruses, and what is an anti-virus (Gole, 2024). Nitesh Kumar Dixit and his coworkers have presented a general overview on computer viruses and defensive techniques. They have observed that computer virus writers commonly use metamorphic techniques to produce viruses that change their internal structure on each infection. On the other hand, anti-virus technologies continually follow the virus tricks and methodologies to overcome their threats (Dixit et al., 2012).

Vivek Kumar has shown that a computer virus is a type of malicious software that spreads between computers and causes damage to data and software. The virus is attached to an executable host file that results in their viral codes executing when a file is opened. The code then spreads from the document or software it is attached to via networks, drives, file sharing programs, or infected email attachment, and can reproduce itself in the same or other form (Kumar, 2024). Hossein Rouhani Zeidanloo and his coauthors have observed that among the diverse forms of malware, botnet and worm are the most widespread and serious threats that occur commonly in most of the cyber-attacks (Zeidanloo et al., 2010).

Milind. J. Joshi and Bhaskar V. Patil have highlighted on the computer virus, history of worst computer attack, and type of computer virus with effect on computer and few examples of virus on their types, working of computer virus, and problem occur due to virus in computers (Joshi & Patil, 2013). Babak Bashari Rad and his coworkers have reviewed new methodologies and have outlined their strengths and weaknesses to encourage those who are interested in more investigation on these areas to make them stronger, more and more, every day (Rad et al., 2011).

3. Research Methodology

Research is the systematic and diligent investigation into a subject to discover, revise, or interpret facts, theories, and applications (Kara, 2012; Mohajan, 2018a). It involves the collection, organization, and analysis of evidence to increase understanding of a topic, characterized by a particular attentiveness to control sources of bias and errors (Roffee & Waling, 2016). The primary purposes of research are documentation, discovery, and interpretation of methods and systems for the advancement of human knowledge (Rocco et al., 2011). Methodology is a systematic and theoretical analysis of the methods and principles associated with a branch of knowledge (Franklin, 2012). It indicates various theoretical commitments about the intended outcomes of the investigation (Howell, 2013). Therefore, research methodology is the systematic and theoretical analysis of the methods applied to a field study, outlining the procedures, tools, and techniques for collecting and analyzing data (Bhome et al., 2013).

We have used secondary data to prepare this qualitative research paper. It can identify the domain, selection, designing and inclusion of various measuring variables in any research. For the collection of secondary data, we have used both published and unpublished data sources related to computer virus. We have unsparingly consulted valuable articles and books of famous authors. We have also collected some materials from the internet and websites (Mohajan, 2017, 2018b, 2020). In this study, we have dealt with some computer virus related qualitative research materials, such as historical background and characteristics of computer virus, various types of computer viruses that are harmful for computer, and protection policy of computer from virus attacks.

4. Objective of the Study

Computer viruses are special kinds of computer programs that spread across disks and networks by making copies of them (Mohajan, 2025b). They are malicious programs that either activate immediately or lie dormant until triggered. They can self-replicate, inserting themselves onto other programs or files, infecting them in the process. They have the ability to get information, check files and documents, hide themselves, and manipulate the programs (Wang et al., 2018). Disasters caused by viruses are damaging of programs and software, deleting of files and data on storage devices, corruption of files, slows down of the speed of the computer, formatting of the hard disk, booting failure, taking up of computer memory, and causing of the system crashes (Kephart & White, 1993). They can write their own codes into the host program, and can infect a computer without permission of the user (Kumar, 2024). At present different unauthorized users hacked our important files, folders and emails through the different worms, viruses, and spam mails (Gole, 2024). Main objective of this article is to discuss the aspects of computer viruses in some detail. Some other minor objectives of the study are as follows:

1) to highlight on the historical background of computer virus,
2) to show the characteristics of computer virus,
3) to analyze the various types of computer viruses, and
4) to discuss the protection of the computer system from virus attacks.

5. Historical Background

A computer virus is a malware that replicates itself by modifying other computer programs and inserting its code. It comes in different types, and each type can affect the devices differently. It is the most well-known form of malware that is designed to disrupt a computer system operation (Aycock, 2006). It results in the loss or damage of hardware, software, data, information, or processing capability of a computer or mobile device. For example, in 1999, the infamous Melissa virus infected thousands of computers and caused damage close to $80 million; while the Code Red worm outbreak in 2001 affected systems running Windows NT and Windows 2000 server and caused damage in excess of $2 billion (Markoff, 1999).

In 1949, Hungarian and American mathematician, physicist, computer scientist, and engineer John von Neumann (1903-1957) has designed a self-reproducing computer program that is considered the first computer virus of the world, and he is considered to be the father of computer virology. His seminal work was published as the "Theory of Self-Reproducing Automata" (von Neumann, 1966). In 1972, Veith Risak has been directly building on von Neumann's work on self-replication published his article "Selbstreproduzierende Automaten mit minimaler Informationsübertragung" (Filiol, 2005). The Creeper virus was first detected on ARPANET, the forerunner of the internet, in the early 1970s that could copy itself. It was an experimental self-replicating program written by Bob Thomas at BBN Technologies in 1971. In the mid-1970s, the "Rabbit" virus emerged that replicated very quickly and caused significant damage at the same pace, but it does not necessarily spread over the network (Parikka, 2007). Rabbit is also called Bacteria, is a malicious code that mainly designed to use up large amounts of system resources, such as message buffers and file space by creating many instances of it or run many times simultaneously (Zeidanloo et al., 2010).

In 1982, Elk Cloner personal computer virus was appeared that was written in 1981 by Richard Skrenta, a ninth grader at Mt. Lebanon High School near Pittsburgh. It spreads through a floppy disk containing a game and attached itself to the Apple II operating system (Jesdanun, 2007). In 1983, the first computer virus was designed and tested by American computer scientist Fred Cohen, and the virus was demonstrated as VAX 11/750 system that crashes the computer within 30 minutes. In 1984, he has written a paper "Computer Viruses: Theory and Experiments" (Cohen, 1984). Two years after the release of Windows 3.0, first virus to specifically targeted Microsoft Windows, WinVir was discovered in April 1992 (Miller, 1996).

The first IBM PC compatible virus in the "wild" was a boot sector virus named Brain or Pakistani virus was created in 1986 and was released in October 1987 by two brothers Amjad Farooq Alvi and Basit Farooq Alvi in Lahore, Pakistan. It struck at the University of Delaware in the USA outside of a test laboratory. It has also struck at many universities of the world and even some businesses, such as the Journal-Bulletin in Rhode Island in the USA (Highland, 1997). One month later, the Lehigh or COMMAND.COM virus was discovered at Lehigh University in Pennsylvania. In December 1987, the Hebrew University at Jerusalem discovered that its microcomputers had been infected by a computer virus in .EXE and/or .COM programs. In 1988, "The Morris" became the first widely spreading virus (Polk et al., 1995).

Upon closer examination the computer specialists under Yisrael Radai of the Computation Center of the University found that every time an .EXE QQAQA executed, and its size increased by 1808 bytes. The .COM programs increased in size as well but there was only a one-time increase. In February 1996, Australian hackers from the virus-writing crew VLAD created the Bizatch virus that was the first known virus to specifically targeted Windows 95 (Yoo, 2004).

6. Characteristics of Computer Virus

A computer virus is a malicious code that replicates and spreads, which is characterized by its ability of attachment to clean files, trigger harmful actions, and hide from antivirus software; and requires user interaction to activate, leading to erratic behavior, crashes, and unauthorized activity (Aycock, 2006). The harmful effects caused by viruses are (Dixit et al., 2012) i) corruption of executable program files, such as word processors, operating system, etc., ii) delete of information from files, iii) increase of the size of files by attaching themselves to those files, iv) monitoring activity and stealing important data, such as credit card information, account number and password, and v) formatting the hard disk for destroying the data (Polk et al., 1995). A computer virus has many characteristics, and some of them are as follows (Mohajan, 2025c):

6.1 Self-Replication Mechanism

The virus has the ability to replicate itself to infect computers, as like its biological counterpart. A self-replicating computer virus is a type of malicious software that upon execution, replicates itself by modifying other computer programs and inserting its own code. It can reproduce and spread automatically (Sayama, 2006). It can copy itself into other executable files, boot sectors, documents with macros, scripts or memory (Kim & Solomon, 2010). The replication may be immediate, delayed, or conditional. By replicating itself it is able to spread across computer systems and networks to infect as much as it possibly can. It requires a host that is executable, document, boot sector, and script to carry and run its code, and some common hosts are .exe/.dll files, Office Macros, PDF/JS, boot records, interpreted scripts, etc. (Wang et al., 2018).

6.2 Executable Path

An executable virus is a non-resident computer virus that stores itself in an executable file and infects other files each time when the file is run. At the starting, the virus must execute itself with infected host to run a program. Its path is not limited to a single location, as it often disguises itself as legitimate files or hides within system folders to evade detection (Yoo, 2004). The virus typically attaches its code to a host file with .exe, .com or .scr extensions and executes by attaching to or replacing legitimate executable files when that file is opened. It can install auto run entries, modify startup scripts, and infect boot sectors to survive reboots (Khan, 2012). It can install itself into Windows system directories to appear legitimate, such as C:\Windows, C:\Windows\System32, or C:\Program Files. Instead of a traditional file path, it targets the hard drive's Master Boot Record (MBR), placing the code at the very beginning of the drive partition to run before the operating system (OS) loads (Arnold & Tesauro, 2000).

6.3 Malicious Payloads

A malicious payload is a code that is responsible for malicious activities and a virus may perform it. It is the component of malware, such as viruses, worms, or Trojans that executes harmful activities, such as file deletion, disk corruption, data wiping, data theft, file destruction, or encryption (Hasan et al., 2020). It ranges from the virus that overwrites or deleted files on the system, such as overwriting the Flash Bios of system or overwriting special media file types to a new type of virus that is not designing for damaging files on the system, but allowing backdoor access to the system for stealing sensitive information, such as passwords (Zeidanloo et al., 2010). The disruptive features are crashes, system instability, and denial of service. The stealthy ways are backdoors, keyloggers, and data exfiltration. Other informational activities are messages, altered UI, etc. The virus lies dormant until triggered, often delivered via phishing, malicious links, or software vulnerabilities (Rajesh et al., 2019).

6.4 Side-Effects

With the self-replicating code, the virus contains a payload that is similar to a warhead on a missile, which is called the side-effect of the virus. The payload has the potential to be malicious, but it does not have to be (Joshi & Patil, 2013). Main side-effects are performance degradation, hijacked browsers, unexpected pop-ups, high resource usage, corrupted files, unauthorized network traffic, disabled security software, unauthorized emails sent to contacts, and compromised confidentiality (Wang et al., 2018). The key side-effects of a computer virus are system performance degradation that makes the computer becomes very slow, freezes often, or takes a long time to open files or programs (Abdulmalik, 2024); data loss or corruption that the files may disappear, become encrypted, or get corrupted, leading to permanent data loss; system instability that frequent crashes, the "blue screen of death", or unexpected shutdowns/restarts (Oyelere & Oyelere, 2015); malicious activities that the computer may display unwanted ads, send unauthorized messages from accounts, or be used for cryptocurrency mining; security disruption that the antivirus software or windows updates may be disabled, preventing fixing the issue (Khan, 2012); hardware malfunction that peripheral devices like printers, keyboards, or mice may behave erratically; and unusual behavior, such that the new, unwanted files, programs, or desktop icons appear on the computer (Joshi & Patil, 2013).

6.5 Disguise

Computer viruses disguise themselves as legitimate files, software, or harmless media to evade detection and trick users into executing them (Mohajan, 2025d). They can disguise themselves before they are noticed by their side-effects. In order to stay far from the anti-virus scanners, computer viruses gradually improve their codes to make them invisible (Rad et al., 2011). There are two methods of disguise: encryption and interrupt interception. The threats of disguise can appear as invoices, reports, or documents, but are actually executable files or malicious containers designed to compromise systems (Aycock, 2006). Encryption is practically the most primitive approach to take cover the operation of the virus code that is the change of the virus body binary codes with some encryption algorithms to hide it from simple view and make it more difficult to analyze and detect. The first encrypting virus, CASCADE, appeared in 1988 (Rad et al., 2012). If the encryption method is failed, the oligomorphic viruses are willing to substitute the decryptor code in new offspring that makes the detection process more difficult for signature based scanning engines. If the encryption and oligomorphism are faced difficulties to protect the anti-virus, polymorphism is developed that has capability to create infinite new decryptors (Szor & Ferrie, 2001). If polymorphism is failed, metamorphic virus is developed to create a new instance, which is not encrypted and does not require decryptor (Jordan, 2002).

7. Types of Computer Viruses

A computer virus is a small piece of code hidden inside a normal program. It can make copies of itself and infect other programs (Mohajan, 2025e). It can damage the system by changing or deleting files and may cause the computer to crash or not work properly (Sarkar, 2018). A virus dropper, usually a Trojan horse, helps put the virus into the computer. When the computer turns on, this virus loads into memory before the operating system starts. It can spread through infected devices like hard drives or USB drives (Wang et al., 2018). We can classify the viruses according to their origin, types of operating system attack, and their techniques of attack, what type of files they infect, how they damage, etc. (Choudhary et al., 2013). There are many types of computer viruses, and in this study, we have discussed mainly nine types of computer viruses as follows:

7.1 Boot Sector Virus

A boot sector virus is a type of virus that infects the part of a storage device and helps to start the computer, called the master boot record (MBR). It is characterized by the fact that it physically resides in the boot sector of the disk rather than in a program file (Kim et al., 2022). It is typically 512 bytes of data on personal computer (PC) that infects the boot sector of a hard disk, a floppy disk, and a pen drive by inserting itself into the boot sector of the disk (Sarkar, 2018). Most of the boot sector consists of a simple and small program that is used to start Disk Operating System (DOS), or whatever operating system is installed. When a computer is booted, the microprocessor first reads the MBR from the disk (Mohajan, 2026a). It is one of the most dangerous viruses, because it infects the MBR, notoriously difficult to remove, and requires a full system format. It is sometimes referred to as memory resident virus because it stays active in the system's memory after execution (Kumar, 2024).

The computer drive has a sector solely responsible for pointing to the operating system so that it can boot into the interface (Mishra, 2012a). The virus is loaded soon after the power on self-test and controls the system through the transferring code and remains in control at all times. It replaces the original boot sector with its own modified version. Once entered it alters the boot sector of the hard disk and remains in the hard disk permanently (Rodionov et al., 2014). It stays in the memory until the system is shut off and the disk is reformatted. Once the virus is loaded, it automatically transfers control to the area where the boot record is available (Aycock, 2006). It overwrites the original boot record with the infected one. It damages or controls the boot sector on the drive, rendering the machine unusable. When other disks are used on the infected computer, the virus is also transferred to their boot sectors (Gole, 2024).

7.2 Spyware

Spyware is a type of malware that is installed on computers and gathers information about the users and stores every keystroke that a user enters without his knowledge. It is designed to monitor computer, and tracks every online and offline activity. It can change web browser's home page or add additional components in the wave page (Kim & Solomon, 2010). It secretly monitors the users' activities and collects information, such as their net surfing habits, types of websites they visit, etc. It records and sends the collected information to advertisers or other interesting parties (Gole, 2024). It is deemed a big danger to the privacy and secrecy of both individuals and organizations. The collected data can also be used to track and capture credit card or bank account information, login and password information or user's personal identity (Wang et al., 2018). Sometimes, it changes certain computer settings which make the computer slow and difficult to work properly. Spyware detection techniques have been presented using three approaches; signature-based, behavior-based, and specification-based (Sheta et al., 2016).

7.3 Macro Virus

A "macro" is a series of instructions, such as menu selections, keystrokes, and commands that are recorded and can be played back with a single command or a keyboard stroke. It is generally triggered by pressing the key or by calling the macro name (Mishra, 2012b). A macro virus is programmed in scripting languages and generally infects document files, such as Microsoft Office Documents, Excel, and similar files without the user's knowledge (Abu & Alta, 2016). It is written in macro languages, such as Visual Basic for Applications (VBA) that is embedded inside a software application, and is relatively easy to learn. Hence, a large number of new macro viruses are created on every day (Singh & Sharma, 2017).

The macro virus resides in macros and use “macro instructions” to perform unintended and sometimes damaging actions. It delivers a payload when the file is opened and the macro runs, and infects data files. It attaches itself to a word processing or a spreadsheet file as a macro (Bontchev, 1996). When an infected document is opened, it can infect all the subsequent documents. It also spreads to other computers on transferring infected documents to those computers. It is easy to create but difficult to detect (Kumar, 2024). In 1999, a macro virus, “Melissa virus” rapidly spread to millions of computer systems worldwide that caused $10 million lost productivity and snarling computer networks with large volumes of email traffic (Mishra, 2012b).

7.4 Polymorphic Virus

A polymorphic virus is a type of malicious software that malware authors hide the malicious code via encryption and can use polymorphic code and signature each time they infect a new system to change the program’s footprint to avoid detection (Gole, 2024). It is very complex virus, and after infecting a system it encrypts itself in a different way. It uses different encryption algorithm each time to make copies of itself, which makes it more difficult for a virus detector to detect, or remove (Choudhary et al., 2013). It has hundreds and thousands of variants. It makes many traditional cyber security tools, such as antivirus or antimalware solutions that rely on signature based detection, and fail to recognize and block the threat (Hamid et al., 2018). It controls the computer, and then decrypts the virus body, and a mutation engine generates randomized decryption routines that changes each time a virus infects a new program. Therefore, it is more difficult for an antivirus to detect and remove them (Tabary, 2011). When polymorphic viruses run, first they decrypt themselves and then behave like any other virus. For example, the “Nuke Encryption Device” (NED) and the “Trident Polymorphic Engine” (TPE) have been written that turn a standard virus into a polymorphic virus (Kumar, 2024). These viruses are the most complex and difficult to detect, often requiring anti-virus companies to spend days or months creating the detection routines needed to catch a single polymorphic. Some polymorphic viruses are Elkern, Marburg, Satan Bug, Tuarey, etc. (Chaumette et al., 2011).

7.5 Multipartite Virus

Multipartite computer viruses infect and spread through multiple methods at the same time, such as executable files and boot sectors, requiring complete removal of all parts for cleanup (Sulianta, 2022). They are fast moving and dual personality viruses and spread in multiple ways at the same time and may create a risk of cyber threats (Hasan et al., 2019). They take different actions on an infected computer depending on variables, such as the operating system installed or the existence of certain files (Sung et al., 2014). These spread across a network by copying themselves or injecting code into critical computer resources. They can simultaneously infect both the boot sector and executable files, allowing them to act quickly and spread fast (Kaur, 2019). Once the PC has been booted from an infected MBR, the virus goes memory resident and infects all accessed, EXE files. If this virus is left undetected, becomes severe threats on the computer system (Kumar, 2024).

7.6 Direct Action Virus

A direct action virus is a type of non-resident computer malware that attaches itself to specific executable files, such as .exe or .com that are specifically designed to perform a destructive activity on system. It can take direct action when a specific condition is met (Salunke & Wadhai, 2022). It can remain dormant until a specific action is taken or a timeframe passes. It stays into the root directory that is located in the hard disk and can take action at the time of booting. It is considered “non-resident”, and it does not install itself or remain hidden in the computer memory (Asaad, 2020). It may be a code authored and can get installed in the memory. When a device user plans to execute a harmless file attached to the malicious code, it instantly delivers a payload. It works by attaching itself to a particular type of file. When someone executes the file, it springs into life, looking for other similar files in the directory for spreading (Choudhary et al., 2013).

7.7 Resident Virus

A resident computer virus is a type of malicious software that installs itself into the RAM of computer rather than just executing and closing, and sit dormant until a payload is delivered. As it resides in memory, it is harder to detect and remove compared to direct-action viruses (Filiol, 2005). It can install itself somewhere in memory, and makes arrangements for the virus body in memory to be executed at some future time (Gajbe, 2020). It is one of the most harmful computer viruses, because it can penetrate any file examined by antivirus tools (Horton & Seberry, 1997).

7.8 Web Scripting Virus

A web scripting virus is malicious software that is designed to breach web browser security and executes harmful code locally on devices (Mohajan, 2026b). This malicious site is created with the purpose to infect code, and this virus is inserted into the sites with the help of infected code and without the knowledge of webmaster. When the virus breaches the web browser security, it injects some malicious codes to take over the web browser and alter some settings (Addison, 1999). It is among the most prevalent threat that users and organizations face. It can capture sensitive information of the device, such as usernames, passwords, and even financial details from unsuspecting users without their knowledge. When any user gives username and password then the interceptor frame hijack all the information about the user. This type of virus is able to propagate faster than any other viruses. There are two types of web scripting virus: persistent and non-persistent web scripting viruses (Szor, 2005). A persistent web scripting virus occurs when malicious code is permanently stored on a website that has the capacity to impersonate the user and cause a lot of damage. A non-persistent web scripting virus occurs when malicious code is not stored on the server. It is often used in phishing schemes and depends on immediate user interaction (Sharp, 2007).

7.9 Browser Hijacker

Browser hijacking is a severe form of cybercrime that affects internet browsers worldwide. It is a malware program that can sneak onto the device through freeware, adware, or spyware. It is the act of altering a user web browser setting without the consent of a user that results in detrimental effects on the privacy and security of a user (Patil & Dash, 2024). It can cause mild disorder to severe threat, and a user can experience inconveniences of facing identity theft that results steal of usernames, passwords, credit cards, and other important information (Zonta & Sathiyanarayanan, 2024). There are a number of browser hijackers, such as Astromenda, Ask Toolbar, Mindspark, Dregol, Binkiland, ESurf, Groovorio, Delta and Claro, Sweet Page, Jamenize, Mazy Search, etc. A browser hijacker may replace the existing home page, error page, or search engine with its own (Singh, 2017). Browser hijackers are developing various trusted brands to cover up their malicious intentions, and such kind of attacks can be avoided by not installing plugins from unknown sources, keeping browser patches up to date, installing antivirus software, taking precautions when downloading software, etc. (Kumar & Indrani, 2018).

7.10 File Infector Virus

A file infector virus is a classic type of malware that attaches itself directly into legitimate executable files, such as .exe, .bat, .dll or .com files and can be dangerous for any device. It can be spread with the help of games and word processors. It is completely self-executive and can make multiple copies of the file itself (Mishra, 2013). It can cause permanent damage to the device. When the infected file is run, the virus becomes active and can spread by infecting other executable files on the system by attaching itself to the end of a file (Morales et al., 2010). It spreads in a device faster, and it is enough to destroy the whole device's workability. It changes the start of a program so that the control jumps to its code (Russell & Gangemi, 1991). It usually transfers control back to the original program, making the infection less noticeable. It is also known as parasitic virus, because it depends on host files to function but keeps the host operational (Kabulov et al., 2020).

7.11 Overwriting Virus

An overwriting virus is a highly destructive malware that targets .exe or .com computer files and changes the original code of an executable program, and then replaces it with malicious code or deletes it (Daoud, 2008). In this situation, the recovering of the original data is almost impossible. It is extremely harmful because it completely destroys elements of a user's system (Maity & Dey, 2021). It is considered as the most harmful virus, because it affects the original code and removes the existing program. Some overwriting viruses are Yankee Doodle, Way, Trj.Reboot, Trivial.88.D, Gorg.202/456, Gorj.377, Loveletter, etc. (Choudhary et al., 2013).

8. Parts of Computer Viruses

A computer virus is a very small portion of a program that can attach itself to another program within the computer, and spreads among computers, and damages data and software. It generally contains three parts: i) the infection mechanism, which finds and infects new files, ii) the payload, which is the malicious code to execute, and iii) the trigger, which determines to activate the payload (Stallings, 2012).

8.1 Infection Mechanism

Computer virus infection mechanisms refer to the processes by which malicious software replicates, hides, and spreads to other files, programs, and computer systems. It is also called the infection vector that has the specific paths or methods, and a computer virus uses these to penetrate a system and spreads its malicious payload through some vulnerability (Şcheau et al., 2015). It differs depending on the used transmission channel and the targets victim. Some viruses have a search routine, which locate and infect files on disk. Other viruses infect files as they run, such as the Jerusalem DOS virus (Filiol, 2005). When the user executes the infected file, the virus activates, replicates itself by modifying other files, and spreads to new systems via networks, shared drives, or email attachments (Zimba, 2017).

8.2 The Trigger

The trigger is also known as a logic bomb, and a time-based trigger is often called a time bomb. It is the part of the virus that determines the condition for which the payload is activated. It is a hidden and unauthorized piece of malicious code that embedded in an application or network (Dusane & Pavithra, 2020). It lies dormant until a specific and predetermined condition is met. It can be positive, when an action occurs, or negative, when an action does not occur. It may happen in a particular date, time, event-based, user actions, size on disk exceeding a threshold, system conditions, presence of other program, and opening a specific file (Bist, 2014). It executes its payload, causing damage to the targeted system when the trigger is activated. The purpose of trigger is to cause harm, disrupt operations, delete or corrupt data, or steal sensitive information (Sharma, 2017).

8.3 Payload

A payload in computing refers to the essential data, message, and files being transmitted through the malicious code executed after a system is breached. It is the body of the virus that executes the malicious activity, such as damaging files, theft of confidential information or spying on the infected system, either intentional or accidental (Farrukh et al., 2022). Accidental damage may result from bugs in the virus, encountering an unknown type of system, or perhaps unanticipated multiple viral infections. Headers and metadata are sent only to enable payload delivery and are considered overhead (Almansour & Mustafa, 2014). Payload activity is sometimes noticeable as it can cause the system to slow down or "freeze". Sometimes payloads are non-destructive and spread a message to as many people as possible, which is called a virus hoax (Wang et al., 2018).

9. Protection System from Virus Attacks

Computer viruses can be easily spread through the emails, email attachments, internet downloads, visiting unknown web pages, installing unknown software, etc. Unwitting users can be victim of hacking easily (Polk et al., 1995). When a computer is infected with viruses, it will show different odd symptoms of its activities, such as it will start to slow down without any reason, abruptly restart itself, often new shortcuts or other icons will be seen in the computer that are not created by the user, visible of unusual error messages and display abnormal behavior, applications installed on the computer will not function as expected (Zhang, 2022). Some files may be deleted of without the permission of the user, etc. It is a good practice to follow some simple precautionary measures that can reduce the possibility of a virus attack (Khan et al., 2017).

Computer virus protection depends on a combination of genuine antivirus software, safe browsing habits, and system updates (Arinze & Agwu, 2024). Anti-virus software is also known as anti-malware and is a computer program that identifies, detects, prevents, inoculates, and takes action to disarm or remove malicious software, such as viruses and worms (Cohen, 1984). It helps users against the latest virus threats. It scans all the files and folders of the computer system to look for the viruses listed in its database of viruses. It removes the virus and repairs the infected file (Zhu et al., 2012).

At present there are a lot of anti-virus software product companies. Some popular antivirus programs are Norton antivirus software, Backdoor antivirus, McAfee virus scan, AVG antivirus, Avast antivirus, Panda antivirus, Dr. Solomon antivirus toolkit, Web scan antivirus, Kaspersky antivirus, Trend Micro antivirus, Avira antivirus, Logic Bomb antivirus, Rabbit antivirus, Bitdefender antivirus, Smadav antivirus, Symantec antivirus, ESET, HTTP (Hyper Text Transfer Protocol), Thunder byte antivirus, Comodo antivirus, Sophos antivirus, Firewall antivirus, F-secure antivirus, USB disk security, Microsoft Security Essentials, McAfee, Malwarebytes, etc. (Coulthard & Vuori, 2002; Filiol et al., 2007).

10. Conclusions

Computer viruses are instructions that are hidden within a computer program and are designed to cause faults or destroy data. They are capable of perpetuating themselves with the basic objective of performing certain activities that could range from annoyance to serious vandalism. They can spread by attaching copies of themselves to some part of a program file, such as a spreadsheet or word processor. In this study, we have studied different types of computer viruses and their characteristics, the threat faced by computer virus, and different preventive measures. If we have the detail knowledge about viruses, we can secure our system very efficiently. The computer viruses are omnipresent as like air in digital world. It passes from computer to computer like a biological virus passes from person to person. But it never occurs naturally and always induced by people. At present the use of computer in digital word is increasing day by day due to the dependency on the networking. Cybercriminals launch and spread virus over the network and steal or alter information for their own benefits. The computer viruses are causing billions of dollars worldwide each year through the economic damages. Various industries are creating, selling or freely distributing antiviruses to users of various operating systems.

References

Abdulmalik, U., I., et al. (2024). A Survey on the Effect of Computer Virus Attacks and Its Preventive Measures among Computer Users. *Kasu Journal of Computer Science, 1*(2), 134-150.

Abu, M. S., & Alta, N. M. (2016). The Rise of Macro Malware in Malaysia. *E-Security, 41*, 6-9.

Addison, C. (1999). Webwatch: Web Security Viruses and Malware. *Information Development, 15*(2), 79-82.

Almansour, S. M, E., & Mustafa, A. B. A. (2014). Effect of Payload Virus as Threatening for Internet Security. *International Journal of Modern Communication Technologies & Research, 2*(9), 39-42.

Arinze, E. D., & Agwu, C. O. (2024). Advancements in Computer Virus Protection: From Origins to Future Trends. *Eurasian Experiment Journal of Scientific and Applied Research, 6*(1), 11-16.

Arnold, W., & Tesauro, G. (2000). Automatically Generated Win32 Heuristic. Virus Bulletin Conference, September 2000.

Asaad, R. R. (2020). Implementation of a Virus with Treatment and Protection Methods. *Icontech International Journal of Surveys, Engineering, Technology, 4*(2), 28-34.

Aycock, J. (2006). *Computer Viruses and Malware*. Publisher: Springer, Canada.

Bhome, S., et al. (2013). *Research Methodology*. Himalaya Publishing House, New Delhi.

Bist, A. S. (2014). Detection of Logic Bombs. (*International Journal of Engineering Sciences & Research, 3*2), 777-779.

Bontchev, V. (1996). Possible Macro Virus Attacks and How to Prevent Them. *Computers & Security, 15*(7), 595-626.

Burger, R. (1991). *Computer Viruses and Data Protection*. Grand Rapids, MI: Abacus.

Chaumette, S., et al. (2011). Automated Extraction of Polymorphic Virus Signatures using Abstract Interpretation. Proceedings of the 5th International Conference on Network and System Security, Milan, Italy.

Choudhary, S., et al. (2013). Awareness about Viruses. *International Journal of Emerging Research in Management &Technology, 2*(4), 42-45.

Cohen, F. (1984). Computer Viruses: Theory and Experiments. *Computers & Security, 6*(1), 22-35.

Coulthard A., & Vuori, T. A. (2002). Computer Viruses: A Quantitative Analysis. *Logistics Information Management, 15*(5/6), 400-409.

Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Method Approaches* (4th Ed.). Thousand Oaks, California: SAGE Publications.

Daoud, E. A. (2008). Computer Virus Strategies and Detection Methods. *International Journal of Open Problems in Computer Science, 1*(2), 122-129.

Dixit, N. K., et al. (2012). The New Age of Computer Virus and Their Detection. *International Journal of Network Security & Its Applications, 4*(3), 79-96.

Dusane, P. S., & Pavithra, Y. (2020). Logic Bomb: An Insider Attack. *International Journal of Advanced Trends in Computer Science and Engineering, 9*(3), 3662-3665.

Farrukh, Y. A., et al. (2022). Payload-Byte: A Tool for Extracting and Labeling Packet Capture Files of Modern Network Intrusion Detection Datasets. TechRxiv.

Filiol, E. (2005). *Computer Viruses: From Theory to Applications*. Springer-Verlag, France.

Filiol, E., et al. (2007). Evaluation Methodology and Theoretical Model for Antiviral Behavioral Detection Strategies. *Journal in Computer Virology, 3*(1), 27-37.

Franklin, M. I. (2012). *Understanding Research: Coping with the Quantitative-Qualitative Divide*. London and New York: Routledge.

Gajbe, A. (2020). Computer Virus: Their Problems & Major Attacks in Real Life. *Journal of Emerging Technologies and Innovative Research, 7*(8), 652-659.

Galvan, J. L. (2015). *Writing Literature Reviews: A Guide for Students of the Social and Behavioral Sciences* (6th Ed.). Pyrczak Publishing.

Gole, D. (2024). Computer Viruses. *National Journal of Hindi & Sanskrit Research, 1*(56), 64-67.

Hamid, I. R., et al. (2018). Classification of Polymorphic Virus Based on Integrated Features. *International Journal n Advanced Science Engineering Information Technology, 8*(6), 2577-2583.

Hannah, S. (2019). Literature Review as a Research Methodology: An Overview and Guidelines. *Journal of Business Research, 104*, 333-339.

Hasan, H., et al. (2020). Enhancing Monkey to Trigger Malicious Payloads in Android Malware. Proceedings of the 17th International ISC Conference on Information Security and Cryptology.

Hasan, M. Z., et al. (2019). Computer Viruses, Attacks, and Security Methods. *LGU Research Journal for Computer Sciences & IT, 3*(3), 20-25.

Highland, H. J. (1997). A History of Computer Viruses: The Famous 'Trio'. *Computers & Security, 16*(5), 416-429.

Horton, J., & Seberry, J. (1997). Computer Viruses: An Introduction. Proceedings of the 20th Australasian Computer Science Conference, Sydney, Australia, February 5-7, 1997.

Howell, K. E. (2013). *Introduction to the Philosophy of Methodology*. London, UK: Sage Publications.

Jesdanun, A. (2007). School Prank Starts 25 Years of Security Woes. Consumer News and Business Channel (CNBC).

Jordan, M. (2002). Dealing with Metamorphism. *Virus Bulletin, 1*(10), 4-6.

Joshi, M. J., & Patil, B. V. (2013). Computer Virus: Their Problems & Major Attacks in Real Life. *International Journal of P2P Network Trends and Technology, 3*(4), 206-209.

Kabulov, A., et al. (2020). Computer Viruses and Virus Protection Problems. *Science and Education, 1*(9), 179-184.

Kara, H. (2012). *Research and Evaluation for Busy Practitioners: A Time-Saving Guide*. Bristol: The Policy Press.

Kaur, J. (2019). Taxonomy of Malware: Virus, Worms and Trojan. *International Journal of Research and Analytical Reviews, 6*(1), 192y-196y.

Kephart, J. O., & White, S. R. (1993). Measuring and Modeling Computer Virus Prevalence. Proceedings 1993 IEEE Computer Society Symposium on Research in Security and Privacy.

Khan, H. A., et al. (2017). Computer Virus and Protection Methods Using Lab Analysis. IEEE 2nd International Conference on Big Data Analysis, pp. 882-885.

Khan, I. (2012). An Introduction to Computer Viruses: Problems and Solutions. *Library Hi Tech News, 29*(7), 8-12.

Kim, D., & Solomon, M. G. (2010). *Fundamentals of Information Systems Security*. Jones & Bartlett Publishers.

Kim, T. N., et al. (2022). Detecting Malware in Portable Executable Files Using Machine Learning Approach. *International Journal of Network Security & Its Applications, 14*(3), 11-17.

Kumar, M. S., & Indrani, B. (2018). A Study on Web Hijacking Techniques and Browser

Attacks. *International Journal of Applied Engineering Research, 13*(5), 2614-2618.

Kumar, V. (2024). Computer-Virus: Types and Its Prevention. *Journal of American Science, 20*(5), 9-12.

Ludwig, M. A. (1996). *The Little Black Book of Computer Viruses: The Basic Technologies* (Volume 1). American Eagle Publications.

Maity, S., & Dey, D. (2021). Computer Virus Attacks. *Pensee International Journal, 51*(3), 585-594.

Markoff, J. (1999). Digital Fingerprints Leave Clues to Creator of Internet Virus. *The New York Times*.

Miller, G. (1996). Technology: 'Boza' Infection of Windows 95 a Boon for Makers of Antivirus Software. *Los Angeles Times*.

Mishra U. (2013). Methods of Repairing Virus Infected Files, A TRIZ based Analysis. *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.2265576

Mishra, U. (2012a). Detecting Boot Sector Viruses: Applying TRIZ to Improve Anti-Virus Programs. *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.1981886

Mishra, U. (2012b). Detecting Macro Viruses: A TRIZ Based Analysis. *SSRN Electronic Journal.* http://dx.doi.org/10.2139/ssrn.1981892

Mohajan, H. K. (2017). Two Criteria for Good Measurements in Research: Validity and Reliability. *Annals of Spiru Haret University Economic Series, 17*(3), 58-82.

Mohajan, H. K. (2018a). *Aspects of Mathematical Economics, Social Choice and Game Theory*. PhD Dissertation. University of Chittagong, Chittagong, Bangladesh.

Mohajan, H. K. (2018b). Qualitative Research Methodology in Social Sciences and Theoretical Economics. *Journal of Economic Development, Environment and People, 7*(1), 23-48.

Mohajan, H. K. (2020). Quantitative Research: A Successful Investigation in Natural and Social Sciences. *Journal of Economic Development, Environment and People, 9*(4), 50-79.

Mohajan, H. K. (2025a). Artificial Intelligence: Prospects and Challenges in Future Progression. *Art and Society, 4*(7), 38-50.

Mohajan, H. K. (2025b). Deep Learning: A Brief Study on Its Architectures and Applications. *Art and Society, 4*(8), 41-49.

Mohajan, H. K. (2025c). Cybercrime: A Potential Threat to Global Community. *Studies in Social Science & Humanities, 4*(6), 28-39.

Mohajan, H. K. (2025d). Vulnerability of Cyber Security is an Unexpected Threat to Global Internet System. *Art and Society, 4*(9), 1-14.

Mohajan, H. K. (2025e). Cyber Child Pornography: An Unexpected Global Heinous Crime against Children. *Art and Society, 4*(10), 7-13.

Mohajan, H. K. (2026a). Addiction and Consumption of Cyber Pornography have Increased Risk Factors and Harmful Effects among Emerging Adults. *Studies in Social Science & Humanities, 5*(1), 33-40.

Mohajan, H. K. (2026b). Machine Learning: A Brief Review for the Beginners. *Innovation in Science and Technology, 5*(1), 26-34.

Morales, J. A., et al. (2010). Identification of File Infecting Viruses through Detection of Self-Reference Replication. *Journal in Computer Virology, 6*(2), 161-180.

Oyelere, S. S., & Oyelere, L. S. (2015). Users' Perception of the Effects of Viruses on Computer Systems: An Empirical Research. *African Journal of Computing & ICT, 8*(1), 121-130.

Parikka, J. (2007). Digital Contagions: A Media Archaeology of Computer Viruses. New York: Peter Lang.

Patil, S., & Dash, C. A. (2024). Effective Detection and Mitigation of Browser Hijacking Using Deep Learning Models. *International Journal of Multidisciplinary Research, 6*(5), 14.

Polk, W. T., et al. (1995). *Antivirus Tools and Techniques for Computer Systems*. William Andrew.

Rad, B. B., et al. (2011). Evolution of Computer Virus Concealment and Anti-Virus Techniques: A Short Survey. *International Journal of Computer Science Issues, 8*(1), 113-121.

Rad, B. B., et al. (2012). Camouflage in Malware: From Encryption to Metamorphism. *International Journal of Computer Science and Network Security, 12*(8), 74-83.

Rajesh, D., et al. (2019). Malwares: Creation and

Avoidance. *International Journal of Computer Sciences and Engineering, 7*(4), 179-183.

Rocco, T. S., et al. (2011). *The Handbook of Scholarly Writing and Publishing*. San Francisco, CA: John Wiley & Sons.

Rodionov, E., et al. (2014). Bootkits: Past, Present & Future. Virus Bulletin Conference, September 2014, pp. 319-326.

Roffee, J., & Waling, A. (2016). Resolving Ethical Challenges When Researching with Minority and Vulnerable Populations: LGBTIQ Victims of Violence, Harassment and Bullying. *Research Ethics, 13*(1), 4-22.

Russell, D., & Gangemi, G. T. (1991). *Computer Security Basics*. O'Reilly.

Salunke, S., & Wadhai, S. A. (2022). Computer Virus and Security. *International Journal of Advanced Research in Computer and Communication Engineering, 11*(4), 754-759.

Sarkar, L. H. (2018). Computer Virus. *International Journal of Creative Research Thoughts, 6*(2), 260-261.

Sayama, H. (2006). Self-Replicating Machines Attempting to Solve the Unsolvable. Workshop on Machine Self-Replication, Tenth International Conference on the Simulation and Synthesis of Living Systems (ALife X).

Şcheau, M. C., et al. (2015). Infection Vectors: Risk Factors for Financial Transactions. *International Journal of Information Security and Cybercrime, 4*(2), 73-85.

Sharma, B. (2017). A Pragmatic Way of Logic Bomb Attack Detection Methodology. *Indian Journal of Science and Technology, 10*(20), 1-5.

Sharp, R. (2007). *An Introduction to Malware*. DTU Library.

Sheta, M. A., et al. (2016). Spyware Detection by Extracting and Selecting Features in Executable Files. Proceedings of the 10th International Conference on Electrical Engineering ICEENG 19-21 April, 2016.

Singh, I., & Sharma, S. (2017). Proposing OMVDPM: Offline Macro Virus Detection and Prevention Mechanism. *International Journal of Computer Applications, 170*(3), 19-22.

Singh, S. (2017). Browser Hijacking: Reasons and Solution. *International Journal for Innovative Research in Multidisciplinary Field, 3*(10), 166-167.

Stallings, W. (2012). *Computer Security: Principles and Practice*. Boston: Pearson.

Sulianta, F. (2022). Comparison of the Computer Viruses from Time to Time. *Central Asia and the Caucasus, 23*(1), 1386-1391.

Sung, P., et al. (2014). Understanding the Propagation Dynamics of Multipartite Computer Virus. *Industrial Management & Data Systems, 114*(1), 86-106.

Szor, P. (2005). *The Art of Computer Virus: Research and Defense*. Publisher: Linda McCarthy.

Szor, P., & Ferrie, P. (2001). Hunting for Metamorphic. *Virus*, 123-143.

Tabary, R. (2011). Context-Free Signatures for Some in-the-wild Polymorphic Viruses. [Online]. Available: www.labri.fr/perso/tabary/publis/polysigns.pdf

Torraco, R. J. (2016). Writing Integrative Literature Reviews: Using the Past and Present to Explore the Future. *Human Resource Development Review, 15*(4), 404-428.

von Neumann, J. (1966). *Theory of Self-Reproducing Automata: Essays on Cellular Automata*. University of Illinois Press.

Wang, J., et al. (2018). The Computer Network Security and Preventive Measures in the Background of Big Data Era. *Advances in Engineering Research*, 137, 413-417.

Yoo, I. (2004). Visualizing Windows Executable Viruses Using Self-Organizing Maps. *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security*, pp. 82-89. 29 October 2004, Washington DC, USA.

Zeidanloo, H. R., et al. (2010). All about Malwares (Malicious Codes). Proceedings of the 2010 International Conference on Security & Management, SAM 2010, July 12-15, 2010, Las Vegas Nevada, USA, 2 Volumes.

Zhang, N. (2022). Computer Virus and Anti-Virus Technology. 3rd International Conference on Language, Art and Cultural Exchange. *Advances in Social Science, Education and Humanities Research*, pp. 91-103, Atlantis Press.

Zhu, Q. et al. (2012). Modeling and Analysis of the Spread of Computer Virus. *Communications in Nonlinear Science and Numerical Simulation, 17*(12), 5117-5124.

Zimba, A. (2017). Malware-Free Intrusion: A Novel Approach to Ransomware Infection Vectors. *International Journal of Computer Science and Information Security, 15*(2), 317-325.

Zonta, T., & Sathiyanarayanan, M. (2024). A Holistic Review on Detection of Malicious Browser Extensions and Links using Deep Learning. In 2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC) (pp. 1-6). IEEE.

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.paradigmpress.org/jpeps
Volume 5 Number 2 June 2026

Design of Thermal Runaway Early Warning and Safety Management System for Polymer Lithium-ion Batteries in Wearable Pet Devices

Haipeng Lin[1]

[1] Shenzhen Jingteng Technology Co., Ltd., Shenzhen 518000, China
Correspondence: Haipeng Lin, Shenzhen Jingteng Technology Co., Ltd., Shenzhen 518000, China.

doi:10.63593/JPEPS.2026.06.07

Abstract

For smart pet collars and health monitors equipped with thin polymer lithium batteries, bending, extrusion and wide temperature swings easily trigger latent thermal runaway, while traditional BMS only supports basic overcharge/over-discharge protection without early detection and graded intervention. This paper analyzes electro-mechanical-environment coupled thermal runaway mechanisms, builds multi-source time-series datasets, and proposes a physics-driven lightweight embedded warning model. A complete perception-edge computing graded safety hardware-software system is developed and validated by finite element simulation, calorimetry and prototype tests. The model reaches 96.4% accuracy for early thermal runaway, issuing alarms 128–176 s earlier than single temperature threshold schemes. The overall system power consumption is only 8.7 mW, enabling staged current limiting and power cut-off to curb thermal runaway evolution. This research provides theoretical and engineering support for battery safety design of pet wearable electronics.

Keywords: wearable pet devices, polymer lithium-ion batteries, thermal runaway, multi-field coupling, lightweight early warning, edge computing

1. Introduction

1.1 Research Background

Smart pet wearables widely adopt 0.8–2.0 mm thin flexible lithium batteries. Long-term skin fitting, repeated mechanical deformation and −10 ℃~50 ℃ humid outdoor environments accelerate diaphragm degradation and internal micro-shorts (Goswami B R, Abdisobbouhi Y, Du H, et al., 2024), which gradually induce thermal runaway, causing device burnout or pet scalds. Commercial products merely rely on single-point temperature and overcurrent protection. Internal hotspots appear long before surface temperature rises, leaving insufficient response time when hazards erupt, which creates obvious safety defects and necessitates targeted research.

1.2 Research Status & Bottlenecks

1) Mechanism: Most thermal runaway studies target large rigid power batteries. Few investigations cover thin cells under cyclic bending; deformation-induced micro-shorts, the primary hazard trigger for pet devices, lack clear evolutionary rules.

2) Warning algorithms: Three mainstream

methods all have drawbacks: physical models demand massive computing resources; fixed thresholds suffer high false alarm rates; deep learning models carry excessive parameters unsuitable for miniature low-power chips. No multi-factor lightweight schemes exist for millimeter-thin batteries.

3) Safety management systems: Conventional wearable BMS lacks tiered early-warning loops, and bulky heat dissipation structures cannot fit compact collars.

Three core bottlenecks remain: ambiguous multi-field coupled failure laws, weak identification of subtle early faults, and safety hardware mismatched with miniature device constraints.

1.3 Research Contents & Three Core Innovations

This paper conducts coupled mechanism analysis, simulation modeling, dataset construction, lightweight model development, integrated safety system design and multi-condition verification.

1) Theoretical innovation: Reveal progressive thermal runaway paths under bending-extrusion and wide-temperature aging, and distinguish failure characteristics of deformation-induced micro-shorts and normal overcharge.
2) Algorithmic innovation: Build a physics-constrained lightweight time-series model integrating temperature, internal resistance, stress and environmental data to capture incipient faults on low-power embedded hardware.
3) Engineering innovation: Develop miniaturized low-power sensing-edge control hardware, formulate four-level closed-loop risk intervention strategies, and complete physical prototype verification.

2. Analysis of Multi-Field Coupling Thermal Runaway Mechanism and Risk Characteristics of Polymer Lithium Batteries in Pet Wearable Devices

2.1 Research Object and Quantification of Typical Working Conditions

This paper selects flexible polymer lithium-ion cells commonly used in pet wearable devices, with a rated voltage of 3.7 V, capacity of 450 mAh and thickness of 1.2 mm, and a minimum allowable bending radius of 15 mm. Basic cell parameters are shown in Table 1.

Table 1. Basic Parameters of Miniature Flexible Polymer Lithium-ion Cells

Parameter	Value
Rated voltage	3.7 V
Charging cut-off voltage	4.2 V
Discharging cut-off voltage	3.0 V
Nominal capacity	450 mAh
Cell thickness	1.2 mm
Operating current range	0–150 mA
Recommended bending radius	≥15 mm

Typical working condition quantification:

1) Mechanical loads: Bending frequency 0–12 times/min, local extrusion stress 0–0.32 MPa, random vibration acceleration 0–2.8 g.
2) Environmental conditions: Ambient temperature −10 °C to 50 °C, relative humidity 25%–85% RH, equivalent heat transfer coefficient of sealed shell heat dissipation 3.2–6.5 $W \cdot m^{-2} \cdot K^{-1}$.
3) Electrical working conditions: Intermittent charge-discharge with an average discharge rate of 0.12 C, floating charge standby accounting for over 70% of total operation time.

2.2 Heat Generation Mechanism of Multi-Field Coupling Thermal Runaway

1) Electrochemical intrinsic heat generation model: The total heat generation rate of the cell consists of polarization heat, reaction heat and side reaction heat, as shown in Formula (1): $Q_{total}=Q_{polar}+Q_{react}+Q_{side}$

variable definition: Q_{total} = total heat generation rate (W); Q_{polar} = polarization heat generation; Q_{react} = reversible heat of electrochemical reactions; Q_{side} = abnormal side reaction heat from SEI decomposition, electrolyte decomposition, negative electrode side reactions, etc. Under normal cycling conditions, the value of Q_{side} remains small; when internal micro-short circuits emerge, side reaction heat rises exponentially and becomes the primary heat source of thermal runaway.

2) Mechanical-thermal coupling triggering mechanism: Repeated bending and

extrusion caused by pet movement lead to internal pole piece dislocation and local plastic deformation of the cell diaphragm, forming micro-pores on the diaphragm and hidden internal micro-short circuits. The contact area of micro-short circuit regions is small, which cannot immediately trigger high-current external short circuits, yet continuous accumulation of local Joule heat forms internal hot spots. Rising temperatures of hot spots further accelerate diaphragm aging and SEI film cracking, forming a positive feedback cycle of "mechanical damage – micro-short circuit – local heat accumulation – material degradation – aggravated short circuit". Such thermal runaway evolves progressively, and the rise of cell surface temperature lags behind internal hot spots, serving as the dominant safety risk source under pet wearable scenarios.

3) Environment-electrothermal coupling evolution law: High ambient temperatures accelerate SEI film decomposition rates and elevate side reaction rates; lithium precipitation occurs at negative electrodes during low-temperature charging, and lithium dendrites pierce the diaphragm to increase micro-short circuit risks; tiny gaps in the shell admit water under high humidity, causing corrosion of cell interfaces and increased leakage current. The sealed wearable shell delivers limited heat dissipation capacity, further amplifying temperature rise effects induced by abnormal heat generation.

4) Thermal runaway stage division & precursor extraction: The entire thermal runaway process of cells under pet wearable working conditions is divided into four stages, with typical observable characteristics of each stage listed in Table 2.

Table 2. Thermal Runaway Evolution Stages and Observable Precursors under Multi-field Coupling

Stage	State Description	Key Observable Signals
Stage I: Safe steady state	Intact cell with no obvious damage	Stable internal resistance, minor temperature fluctuations, no abnormal temperature rise rate
Stage II: Early abnormal budding	Slight diaphragm micro-damage and faint micro-short circuits with no obvious external signs	Slight rise in internal resistance, increased local temperature difference, accumulated continuous stress, surface temperature rise rate <0.15 °C/min
Stage III: Thermal runaway development	Aggravated side reactions and expanding hot spots	Voltage fluctuations, significantly elevated internal resistance, temperature rise rate 0.15–1.0 °C/min
Stage IV: Thermal runaway eruption	Diaphragm melting, violent side reactions, gas generation and ignition	Sharp voltage drop, temperature rise rate >1 °C/min, cell bulging and gas emission

Stage II is the golden window for early warning. At this stage, external voltage and overall temperature change faintly, which cannot be identified by traditional single-point temperature protection, requiring joint judgment of multi-source signals including internal resistance, temperature difference and stress.

2.3 Finite Element Simulation Modeling of Multi-field Coupling

- **Modeling settings**: Establish an electrochemical-temperature-stress multi-physics coupling simulation model geometrically built at a 1:1 scale for 1.2 mm thick flexible cells. Cyclic bending loads are applied to mechanical boundaries; the electrochemical module couples the intrinsic battery heat generation equation; the heat transfer module considers sealed heat dissipation boundaries of wearable shells.

- **Simulation objectives**: Solve the internal temperature field and stress field distribution of cells under different bending stress conditions, and obtain the evolution process of internal hot spots. Simulation results show that when local extrusion stress exceeds 0.27 MPa and the equivalent plastic strain of local diaphragms surpasses 0.08, millimeter-level local hot spots begin to form inside cells, with hot spot temperatures 12–21 °C higher than cell surface temperatures. This intuitively explains the physical phenomenon that "internal hot spots appear prior to surface temperature rise". Datasets output from simulations compensate for the shortcomings of scarce and high-cost real thermal runaway test samples, providing partial sample support for subsequent training of early warning models.

3. Construction of Lightweight Intelligent Thermal Runaway Early Warning Model Based on Multi-sensor Fusion

3.1 Design Constraint Indicators of Early Warning Model

For embedded hardware of pet wearable devices, the model must balance identification accuracy and embedded computing power constraints, with key indicators shown in Table 3.

Table 3. Design Indicators of Early Warning Model

Indicator	Target Value
Identification accuracy for Stage II thermal runaway	≥95%
Single sample inference time	< 12 ms
Model parameter quantity	< 12 k
Power consumption of single-cycle inference	< 2.2 mW
False alarm rate	≤3.0%

3.2 Feature Set Construction and Data Preprocessing

3.2.1 Multi-dimensional Input Feature Set

Abandoning the traditional single temperature threshold scheme, a five-category time-series feature set is constructed:

- Electrical features: Terminal voltage U_t, charge-discharge current I_t, AC internal resistance R_{ac}, voltage fluctuation rate $\Delta U/\Delta t$.
- Temperature features: Multi-point cell surface temperatures T_1,T_2,T_3, maximum temperature difference ΔT_{max}, temperature rise rate dT/dt.
- Mechanical features: Real-time extrusion stress σ, bending cycle count.
- Environmental features: Ambient temperature T_{amb}, ambient relative humidity RH.
- Aging features: Cycle count, capacity attenuation coefficient.

3.2.2 Data Collection and Preprocessing

A working condition simulation test platform is built to conduct cyclic bending, high-low temperature cycling, accelerated aging and micro-short circuit induction tests on cells, collecting a total of 42,600 groups of time-series samples including 3,820 groups of abnormal samples (Shi H, Xu Y, Geng J, et al., 2026). Raw data contains noise and asynchronous timing issues, processed sequentially via outlier elimination, sliding window filtering, multi-sensor time alignment and min-max normalization. The Pearson correlation coefficient is adopted for feature screening to eliminate redundant features with a correlation coefficient lower than 0.3, finally retaining 11-dimensional core input features to reduce model input dimensions. The calculation formula of the Pearson correlation coefficient is shown in Formula (2):

$$r_{xy}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

3.3 Architecture of Lightweight Time-Series Early Warning Model Integrated with Physical Information

3.3.1 Overall Model Architecture

1) The model adopts a hybrid architecture of "physical prior constraint + lightweight time-series network". Battery thermal

dynamics equations are embedded into the loss function as physical constraint terms to restrict model outputs to comply with battery heat generation physical laws, reducing the risk of generalization failure of pure data-driven models outside working conditions.

2) The main network adopts a lightweight time-series attention mechanism, assigning higher attention weights to time-series fragments with faint early abnormalities and strengthening extraction capacity of budding-stage features.
3) A risk judgment module is connected after the network to output a risk probability value ranging from 0 to 1, mapped to four risk levels.

Loss function definition, as shown in Formula (3): $L_{total}=L_{cls}+\lambda L_{phy}$ parameter explanation: L_{cls} = classification cross-entropy loss; L_{phy} = physical constraint loss; λ = weight coefficient of physical constraint, set as $\lambda=0.25$ in this paper.

3.3.2 Graded Early Warning Judgment Logic

Risk grading standard: Four risk states are divided according to the model output risk probability P_r:

1) Safe (Green): $P_r<0.30$, normal cell status.
2) Level-I Early Warning (Yellow): $0.30\leq P_r<0.55$, entering the early abnormal budding stage.
3) Level-II Alarm (Orange): $0.55\leq P_r<0.80$, continuous deterioration of thermal abnormalities.
4) Level-III Emergency Alarm (Red): $P_r\geq 0.80$, imminent thermal runaway.

System linkage rule: Different risk levels directly correspond to differentiated intervention strategies of the back-end safety management system, realizing linkage between "early warning and control".

3.4 Model Training and Comparative Evaluation

To fully verify the comprehensive performance, generalization ability and engineering adaptability of the proposed lightweight early warning model, iterative training and hyperparameter optimization are carried out based on the preprocessed multi-source time-series dataset. Multiple groups of mainstream classic early warning algorithms are set as control groups for horizontal comparative tests, with quantitative evaluation conducted from multi-dimensions including identification accuracy, inference efficiency and early warning timeliness.

- Dataset partition: The dataset is divided into training set, verification set and test set at a ratio of 7:2:1. After hyperparameter tuning, the final model parameter quantity reaches 8.4 k, satisfying embedded deployment constraints.
- Control group setup: To highlight the superiority of the proposed model in early thermal runaway early warning scenarios, the industrially adopted fixed temperature threshold early warning algorithm and traditional standard LSTM time-series prediction model are selected as control groups for horizontal comparison tests under identical test datasets and hardware environments. A comparison of core performance indicators of each model is shown in Table 4.

Table 4. Performance Comparison of Different Early Warning Schemes

Model Scheme	**Accuracy /%**	**Recall Rate /%**	**False Alarm Rate /%**	**Inference Time /ms**	**Advanced Warning Duration /s**
Fixed temperature threshold	74.2	61.3	11.7	<1	42
Standard LSTM network	94.7	92.1	3.8	47	113
Proposed physical information lightweight model	96.4	94.6	2.7	8.6	151

The traditional temperature threshold scheme delivers short advanced warning duration and low accuracy; the standard LSTM achieves favorable identification effects yet suffers long inference time, making it unsuitable for miniature wearable devices. The proposed model balances identification accuracy and inference speed, with significantly longer advanced warning duration than control schemes, capable of identifying faint abnormalities in the budding stage of thermal runaway.

4. Overall Hardware & Software Design of Battery Thermal Safety System for Pet Wearables

This chapter designs an integrated battery safety system targeting thin lithium batteries' risks (bending, wide temperature aging, hidden micro-shorts). It builds a full "sensing-inference-decision-protection-alarm" workflow for early battery fault detection and tiered protection, optimized for tiny, low-power wearable hardware.

4.1 System Framework & Design Rules

Hierarchical Structure

- Sensing layer: Collect 6 signals (3-channel cell temperature, voltage/current, thin-film stress) as model inputs.
- Edge computing layer: Local low-power RISC-V MCU runs signal filtering, lightweight model inference and risk judgment; one inference cycle takes ~12 ms.
- Interaction layer: Onboard tri-color LEDs show real status; BLE pushes alerts to phones; cloud only stores 5-min interval logs without real-time calculation.

Core Design Constraints

Limited inner space (220 mm²) and standby current (142 μA) set five rules: miniaturized sensing PCB (210 mm²), ultra-low power with adjustable 1–10 s sampling, full local risk computation, four-tier closed-loop control, wide-temperature components (-10 ℃ ~ 50 ℃) supporting 6 bending cycles per minute.

4.2 Modular Hardware Design

The hardware adopts split independent modules for compact layout and low power draw:

1) Multi-source sensing module: Three NTC temperature probes eliminate single-point measurement bias; thin-film stress sensor captures extrusion & bending deformation.
2) Edge MCU module: 32-bit low-power RISC-V chip with enough on-chip RAM to run the lightweight model without extra storage; handles sensor reading, data preprocessing and risk classification.
3) Dual electrical protection module: Software current limiting + independent hardware cut-off circuits. Hardware instantly disconnects charge/discharge loops at Level-III alarms to avoid software failure risks. Thermal buffer layers reduce impact and heat spread.
4) BLE human-machine module: Bluetooth syncs battery health and warnings to mobile APP; tri-color LEDs correspond to four risk levels for offline status viewing.

Power Test Results

The full prototype consumes 8.7 mW during operation and 142 μA in sleep mode; the safety system barely reduces device battery life.

4.3 Embedded Software & Tiered Protection Logic

The embedded software on edge MCU handles data processing, fault prediction and dynamic protection, split into six decoupled sub-modules: bottom driver, data preprocessing, risk inference, tiered strategy scheduler, Bluetooth communication, battery health evaluation.

Operation Flow

Periodic multi-sensor data → filtering & normalization → local lightweight model outputs risk grade → system executes matched control, saves logs and sends mobile alerts when needed.

Four-Tier Closed-Loop Control

1) Safe (Normal): Standard charge-discharge management, continuous monitoring and real-time SOH update.
2) Yellow Level-I Warning (Minor anomaly): Lower charging current, disable idle peripherals, log data silently without mobile notifications.
3) Orange Level-II Alarm (Worsening fault): Disable fast charging, cap output power; push APP reminders for equipment inspection.
4) Red Level-III Emergency (Thermal runaway risk): Hardware cuts power entirely, device locks; APP sends high-risk alerts to stop usage immediately.

5. System Test Verification and Result Analysis

To verify the early warning capacity, protection functions and environmental adaptability of the battery thermal safety management system, a test platform is built, and control groups and experimental groups are set for multi-dimensional comparative testing. Verification is completed from three layers: model algorithm, whole machine reliability and physical prototype, followed by analysis of test results and existing deficiencies.

5.1 Test Platform and Test Scheme

The test platform mainly includes a high-low temperature environmental test chamber, dynamic bending-extrusion testing machine, accelerated rate calorimeter (ARC), multi-channel battery test system and power analyzer.

- Control group: Adopt original BMS of commercial pet wearable devices, only supporting single-point temperature monitoring and basic overcharge-overdischarge protection.
- Experimental group: Equipped with the complete early warning and safety management system designed in this paper.

5.2 Special Verification of Early Warning Model

Multiple groups of progressive micro-short-circuit thermal anomalies are induced under different ambient temperature, bending stress and cell aging conditions, and early-warning trigger moments are recorded. Test results indicate that for deformation-induced progressive thermal runaway, this system achieves an average early-warning lead time of 151 s, with a range of 128-176 s (Li H G, Zhou D, Zhang M H, et al., 2022); the single-point temperature protection of the control group mostly responds at the late stage of thermal runaway, with an average of merely 44 s. Under high-low temperature conditions of -5 ℃ and 45 ℃, the model recognition accuracy is higher than 93%, the false-alarm rate is controlled within 3.0%, and multi-source feature input can effectively reduce misjudgment caused by environmental disturbance.

5.3 Whole Machine System Function and Reliability Test

1) Power consumption & function test: Power analyzer measurement shows the whole machine operating power consumption is 8.7 mW, consistent with design indicators. Simulated four-level risk scenarios are repeated 50 times, with graded alarms, current limiting and power-off protection triggered normally without missed or false triggering.
2) Environmental reliability test: A total of 120 h temperature cycle impact tests from −10 °C to 50 °C are completed, alongside cyclic bending durability tests at 6 times per minute. No failure occurs in sensor collection, model inference and protection functions after testing, proving the system's adaptability to complex pet wearable working conditions.

5.4 Physical Loading Test

The system is integrated into a pet smart collar prototype for a 30-day simulated wearing test reproducing pet movement and diurnal outdoor temperature variations, using a total of 8 prototypes (4 for the experimental group and 4 for the control group). Three Level-I early warnings emerge in the experimental group induced by slight internal resistance rise from accumulated battery stress. The system automatically limits charging current without safety accidents. Two local overheating phenomena occur in the control group, with protection activated only after a sharp temperature rise and lacking capacity for early intervention. Endurance tests show that equipment endurance declines by 6.2% after installation of the safety system, an acceptable margin. (Zhang J R & Xuan F Z, 2026)

5.5 Discussion of Test Results

Tests prove that relying solely on single temperature parameters fails to identify early internal micro-short circuits induced by deformation. The proposed system fuses multi-source features including internal resistance, stress and temperature difference, capturing faint abnormal cell characteristics via the lightweight model to realize early warning in the budding stage of thermal runaway. The system still has certain limitations: the fault evolution speed of instantaneous severe external short circuits leaves insufficient identification windows for algorithms, making system safety protection highly dependent on hardware backup protection. Individual cell differences under mass production working conditions cause baseline offset of monitoring features. Online self-calibration logic can be optimized in follow-up research to further improve model robustness.

6. Conclusions and Prospects

6.1 Main Conclusions

Repeated bending and extrusion under pet wearable scenarios cause diaphragm micro-damage inducing internal micro-short circuits, serving as the primary inducement of thermal runaway of miniature polymer lithium batteries with progressive evolution characteristics. Internal hot spots emerge earlier than cell surface temperature rise, leading to obvious hysteresis in traditional single-point temperature protection.

The lightweight time-series early warning model integrated with physical information constructed in this paper contains only 8.4 k parameters, achieving a test set identification accuracy of 96.4%. It significantly advances thermal runaway alarm time compared with traditional schemes and adapts to embedded low-computing-power hardware.

The designed integrated multi-source perception-edge control safety management system delivers an average hardware power consumption of 8.7 mW and establishes a four-level risk closed-loop graded intervention strategy (Mussa A S, Klett M & Lindbergh G, 2024), realizing complete protection covering early current limiting regulation to hardware power-off under high-risk states. Physical loading tests verify that the system can effectively identify and suppress thermal runaway risks with limited impact on endurance, suitable for pet wearable devices.

6.2 Research Deficiencies

Under extreme instantaneous short-circuit working conditions, fault signals leave limited identification windows for algorithms, and system safety protection relies heavily on hardware backup protection. Meanwhile, individual differences of cells from different batches induce baseline offset of monitoring features. In addition, after long-term aging cycles of batteries, the online self-calibration capacity of model parameters in the current framework remains to be further improved, increasing the difficulty of battery fault identification and risk early warning implementation for smart pet wearable devices.

6.3 Future Prospects

Follow-up research can be carried out in the following directions: adding online adaptive calibration algorithms for cells oriented to mass-produced products; introducing digital twin technology into full-life-cycle health management of miniature batteries; exploring integrated integration schemes of flexible sensors and cells; further compressing model scale to expand application to lower-power miniature wearable terminals.

References

Goswami B R, Abdisobbouhi Y, Du H, et al. (2024). Advancing battery safety: Integrating multiphysics and machine learning for thermal runaway prediction in lithium-ion battery module. *Journal of Power Sources, 589*, 233782. DOI:10.1016/j.jpowsour.2024.233782.

Li H G, Zhou D, Zhang M H, et al. (2022). Multi-field interpretation of internal short circuit and thermal runaway behavior for lithium-ion batteries under mechanical abuse. *Energy, 263*, 126027. DOI:10.1016/j.energy.2022.126027.

Mussa A S, Klett M, Lindbergh G. (2024). Design parameters affecting mechanical failure and electrochemical degradation of ultrathin Li-ion pouch cells under repeated flexing. *Frontiers in Batteries and Electrochemistry, 1*, 1371167. DOI: 10.3389/fbael.2024.1371167

Shi H, Xu Y, Geng J, et al. (2026). Data-Driven Multi-Sensor Early Warning and Risk-Oriented Prognostics for Lithium-Ion Battery Thermal Runaway. *Electronics, 15*(13), 2909. 1-23. DOI:10.3390/electronics15132909.

Zhang J R, Xuan F Z. (2026). A Self-Decoupling Multimodal Sensor for Enhanced Early Warning of Lithium-Ion Battery Thermal Runaway. *Research, 9*, 1120. DOI:10.34133/research.1120.

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.paradigmpress.org/jpeps
Volume 5 Number 2 June 2026

Online Monitoring and Performance Degradation Evaluation Method for Automotive-Grade RF Power Amplifiers Under Endurance Aging

Yanbo Zhao[1]

[1] Shanghai Lixun Electronic Technology Co., Ltd., Hangzhou 310012, China
Correspondence: Yanbo Zhao, Shanghai Lixun Electronic Technology Co., Ltd., Hangzhou 310012, China.

doi:10.63593/JPEPS.2026.06.08

Abstract

To meet AEC-Q102 reliability standards for automotive RF power amplifiers (PAs), traditional offline aging tests suffer from discontinuous data, inability to distinguish reversible thermal drift from permanent device damage, and oversimplified evaluation metrics. This paper proposes a complete endurance assessment system for automotive PAs. We build a multi-stress coupled online monitoring platform to continuously collect multi-dimensional performance parameters. A normalized Comprehensive Performance Degradation Index (CPDI) integrating failure physics and time-series features is proposed to quantify device degradation. A hybrid FP-LSTM prediction model constrained by failure physics is put forward to address weak generalization and low precision of conventional algorithms. A four-level hierarchical evaluation framework complying with AEC-Q102 and ISO 26262 is also formulated. Experiments on 24 GaAs/GaN PAs under 1000 h composite aging verify that the platform's maximum measurement error is below 2.08% and can detect early hidden degradation. CPDI achieves 98.6% accuracy in identifying irreversible damage. The FP-LSTM model reduces lifespan prediction mean relative error to only 3.17%, boosting accuracy by 47.52% versus the classic Arrhenius model. The proposed system can be deployed on production lines to provide standardized quantitative support for automotive RF PA endurance certification.

Keywords: automotive-grade RF power amplifier, accelerated aging, online monitoring, degradation quantification, failure physics, FP-LSTM, AEC-Q102

1. Introduction

1.1 Research Background and Engineering Demand

Vehicle millimeter-wave radar and 5G/V2X RF PAs require stable 10–15 year service life. Actual vehicle conditions superimpose temperature cycling, humidity bias, high-power RF excitation and mechanical vibration, triggering micro-failures such as electromigration and gate dielectric damage, which degrade gain, efficiency and linearity and may lead to communication/radar breakdown violating ISO 26262. AEC-Q102 only defines aging stress limits without standardized online monitoring and quantitative evaluation schemes. Widely used offline testing yields fragmented data, misses early degradation signs, and cannot separate thermal drift from permanent damage. Judging failure merely by a 1 dB S21 drop causes frequent misjudgments. This paper develops an integrated evaluation framework covering online monitoring, degradation quantification, life prediction and hierarchical grading to satisfy automotive PA endurance certification demands.

1.2 Research Status and Existing Gaps

1) **Aging monitoring**: Existing solutions are either offline discrete testing or single-DC online acquisition. They cannot synchronously capture full RF and distortion indicators and must halt aging during measurement; no multi-dimensional continuous online monitoring under multi-stress coupling exists.
2) **Degradation life prediction**: Traditional physical models like Arrhenius only fit single-stress environments with poor multi-stress precision. Pure data-driven LSTM/BP lack physical constraints, overfit small aging datasets and generalize poorly across different PA processes. No hybrid physics-data model targets automotive RF PAs.
3) **Endurance evaluation system**: The industry lacks unified quantitative evaluation criteria. The single S21 1 dB threshold is error-prone due to up to 0.8 dB low-temperature reversible drift and ignores efficiency, leakage and linearity degradation, failing AEC-Q102 and ISO 26262 quantitative assessment requirements.

1.3 Three Core Innovations

1) **Hardware platform**: A nanosecond synchronous RF switching structure realizes multi-parameter synchronous sampling without cutting off high-power aging excitation. Monitored items cover DC characteristics, S-parameters, saturation power, PAE and harmonic distortion, with overall measurement error ≤2.08%.
2) **Quantification & prediction model**: A 0–1 normalized CPDI is constructed via failure mechanism weighting and grey correlation analysis to comprehensively quantify multi-dimensional degradation. The FP-LSTM hybrid model embeds multi-stress physical degradation equations as prior constraints to suppress small-sample overfitting, lifting prediction accuracy by 47.52% against traditional physical models.
3) **Evaluation system**: Calibrated four-tier degradation thresholds plus a thermal drift filtering algorithm are established. A full standardized online evaluation process compliant with AEC-Q102 and ISO 26262 supports early latent damage detection and root failure mechanism backtracking.

2. Degradation Mechanisms of Automotive-Grade RF PAs Under Coupled Multi-Stress and Screening of Characteristic Parameters

2.1 Device Structures of Automotive PAs and Decomposition of Composite Service Stresses

Two mainstream types of automotive-grade RF PAs are investigated in this paper: GaAs pHEMT amplifiers for 24 GHz short-range radar and GaN HEMT high-power amplifiers for 77 GHz main radar. The core transistor structure consists of four critical failure-prone regions: gate dielectric layer, channel active region, source-drain metal interconnections, and passivation protective layer. In accordance with AEC-Q102 and ISO 16750 automotive electrical standards, composite service stresses acting on devices are decomposed into four independent coupled loads, as summarized in Table 1.

Table 1. Boundary Conditions of Composite Service Stresses for Automotive PAs

Stress Type	Parameter Range	Loading Mode	Dominant Micro-Failure Modes
Temperature cycling stress	−40 ℃ ~ 125 ℃, temperature ramp rate of 5 ℃/min	Periodic circulation	Hot Carrier Injection (HCI), thermal fatigue
Temperature-humidity bias stress	85 ℃, 85% RH, continuous DC bias applied	Long-term static loading	Gate metal corrosion, water vapor penetration of passivation layer
High-power RF excitation	Continuous-wave output power of 18~32 dBm, 77 GHz modulated signals	Uninterrupted RF output	Load-pull thermal fluctuation, gate dielectric fatigue

Triaxial mechanical vibration	Random vibration of 10~2000 Hz, maximum acceleration of 10 g	Continuous low-amplitude vibration	Solder layer microcracks, matching impedance offset

The four types of stresses do not act independently. High temperatures accelerate electromigration rates, while microcracks induced by vibration elevate device thermal resistance and further amplify junction temperature fluctuations, forming a positive feedback loop that accelerates degradation.

2.2 Microscopic Failure Physics Mechanisms Under Coupled Multi-Stress

2.2.1 Hot Carrier Injection (HCI) Degradation Mechanism

Under alternating high-low temperature cycling combined with continuous high-power RF excitation, high-speed electrons inside channels undergo impact ionization. High-energy carriers break down gate dielectric interfaces and accumulate fixed interface state charges, which reduce transistor transconductance g_m, directly leading to attenuation of small-signal S21 gain and drift of static drain current I_{DQ}. HCI corresponds to irreversible micro-damage; RF performance cannot recover after stress removal, rendering it the primary long-term endurance degradation mode of automotive PAs. The HCI degradation rate follows the Arrhenius temperature acceleration model as shown in Equation (1): $k_{HCI} = A_{HCI} \exp\left(-\frac{E_{a,HCI}}{k_B T_j}\right)$.

where k_{HCI} denotes the HCI degradation rate; A_{HCI} is a process-related pre-exponential coefficient; $E_{a,HCI}$ represents the HCI activation energy; k_B is the Boltzmann constant; T_j refers to real-time transistor junction temperature.

2.2.2 Metal Interconnection Electromigration (EM) Degradation Mechanism

Under long-term constant current and high-temperature loading, metal atoms along source-drain interconnections migrate along the current direction, generating voids and protrusion defects that continuously raise contact resistance. Electromigration directly reduces PAE, increases static power consumption, and further elevates junction temperature to form a positive feedback cycle accelerating degradation. The inverse power-law acceleration model for electromigration mean time to failure is shown in Equation (2):

$$k_{\mathrm{HCI}} = A_{\mathrm{HCI}} \exp\left(-\frac{E_{a,\mathrm{HCI}}}{k_B T_j}\right)$$

where $MTTF_{EM}$ is the mean time to electromigration failure; J denotes metal current density; n is the current acceleration exponent; $E_{a,EM}$ represents the electromigration activation energy. The electromigration degradation inverse power-law acceleration model is as follows:

$$MTTF_{EM} = A_{EM} J^{-n} \exp\left(\frac{E_{a,EM}}{k_B T_j}\right)$$

2.2.3 Gate Corrosion Degradation Mechanism Under Coupled Temperature and Humidity

Water vapor penetrates passivation layers under high-humidity environments and contacts gate metal. Electrochemical corrosion occurs under DC bias, causing exponential growth of gate leakage current I_G and severe deterioration of total harmonic distortion (THD) and third-order intermodulation distortion (IMD3). This failure mode accounts for over 30% of device failures in automotive equipment deployed in humid southern regions.

2.2.4 Vibration-Assisted Impedance Offset Degradation Mechanism

Continuous triaxial vibration initiates microcracks in package solder layers, shifting input and output matching network impedances and deteriorating input reflection coefficient S11 and output reflection coefficient S22 (Zhang H, Liu Y & Wang L., 2023). RF power transmission efficiency declines, and load-pull fluctuations introduce periodic stress damage to gate dielectrics, indirectly accelerating HCI and electromigration processes.

2.3 Screening of Degradation Characteristic Parameters and Criterion for Distinguishing Reversible and Irreversible Degradation

High-sensitivity monitoring parameters are screened across four dimensions (DC static, small-signal RF, large-signal nonlinear, thermal state) based on the four failure mechanisms, as listed in Table 2. Two types of degradation variation rules are clarified: parameter offsets induced by instantaneous temperature fluctuations are defined as reversible thermal drift, which fully recovers after stress removal and temperature returns to ambient conditions. Monotonic parameter shifts caused by accumulated interface states, metal corrosion, and solder layer microcracks correspond to permanent irreversible damage, which constitutes the core evaluation target for endurance testing.

Table 2. Classification of High-Sensitivity Degradation Monitoring Characteristic Parameters

Parameter Dimension	**Core Monitoring Indicators**	**Dominant Corresponding Failure Modes**	**Degradation Reversibility**
DC static parameters	I_{DQ} , gate leakage current I_G , transconductance g_m	HCI, electromigration, gate corrosion	Primarily permanent irreversible
Small-signal RF parameters	S21 gain, S11, S22 reflection coefficients	Vibration-induced impedance offset, HCI	Primarily permanent irreversible
Large-signal nonlinear parameters	Saturation power P_{sat} , 1 dB compression power P_{1dB} , PAE, THD, IMD3	Gate dielectric fatigue, electromigration	Primarily permanent irreversible
Thermal state parameters	Real-time junction temperature T_j , steady-state temperature rise fluctuation amplitude	All coupled stresses	Short-term reversible fluctuation

3. Design of Online Synchronous Monitoring Platform for Endurance Aging Under Coupled Multi-Stress

3.1 Overall Layered Architecture of the Platform

The proposed online monitoring platform consists of four hardware layers: multi-stress coupled environmental loading unit, RF time-sharing synchronous acquisition unit, lower computer real-time control and caching unit, and upper computer data preprocessing and storage unit. The hardware topology adopts time-sharing RF switch multiplexing design, enabling synchronous multi-parameter collection without cutting off high-power aging excitation of PAs, thus guaranteeing uninterrupted application of composite aging stresses. Key hardware specifications are listed in Table 3.

Table 3. Key Hardware Equipment Specifications of the Online Monitoring Platform

Equipment Module	**Model**	**Critical Technical Indicators**
Temperature-humidity cycling chamber	TH-1000	Temperature: −40 ℃ ~ 130 ℃; humidity: 10%–95% RH
Triaxial vibration test bench	EV-300	Frequency range: 10–2000 Hz, maximum acceleration: 20 g
RF signal generator	Keysight N5183B	Output of 24/77 GHz millimeter-wave modulated signals
Embedded vector network analyzer (VNA)	Keysight N4431A	Measurement error ≤ 0.12 dB, nanosecond-level synchronization for time-sharing switching
Programmable DC source	Keithley 2400	I_G/I_{DQ} acquisition resolution: 1 μA
High-speed harmonic analyzer	FSW43	Dynamic test range of 60 dB for THD and IMD3

Embedded control unit	STM32H743	Maximum synchronization delay of multi-parameter acquisition ≤ 500 ns

3.2 Timing Control Strategy for Time-Sharing Synchronous RF Acquisition

Conflicts exist between high-power continuous aging excitation and VNA small-signal test RF paths. A nanosecond-level single-pole double-throw RF switch timing synchronization logic is developed. The complete cycle T_{cycle}=10 ms is divided into two phases:

1) **Continuous aging excitation phase (9.9 ms):** RF switches connect to power load paths, and standard automotive high-power RF excitation is continuously applied to PAs with uninterrupted composite temperature-humidity-vibration stresses.
2) **Synchronous acquisition phase (0.1 ms):** RF switches rapidly toggle to VNA and harmonic acquisition paths to synchronously collect S-parameters, output power, distortion, DC current, and substrate temperature data, then switch back to aging paths immediately after measurement completion.

The full timing logic is uniformly triggered by the STM32 embedded chip, with a maximum synchronization delay of 486 ns across all measuring devices, eliminating parameter matching errors caused by timing misalignment.

3.3 RF Link System Error Calibration and Junction Temperature Compensation

3.3.1 RF Full-Link SOLT Calibration

Cables, RF switches, and power dividers introduce system measurement errors due to temperature variations in their attenuation coefficients. The SOLT full-band calibration method is employed to calibrate the RF path from 24 to 77 GHz at each temperature point, establishing a temperature-attenuation error correction lookup table model. The host computer collects data and performs real-time RF parameter compensation to eliminate system drift caused by temperature changes.

3.3.2 Real-Time Junction Temperature Accurate Measurement Model

The temperature probe cannot be directly attached to the active region of the transistor. This paper employs a power dissipation-thermal resistance model to indirectly calculate the junction temperature, with the calculation formula shown in Equation (3):

$$T_j = T_{\text{sub}} + P_{\text{diss}} \cdot R_{\text{th,jc}}$$

where T_{sub} denotes measured PA substrate temperature; P_{diss} represents real-time device power dissipation; $R_{th,jc}$ is steady-state chip-to-package thermal resistance calibrated from manufacturer process datasheets. The junction temperature calculation formula is as follows:

$$T_j=T_{sub}+P_{diss}\cdot R_{\mathrm{th},jc}$$

3.4 Calibration Test of Platform Acquisition Accuracy

Standard reference GaN PA samples are tested simultaneously using the proposed online monitoring platform and high-precision offline laboratory instruments under 10 steady-state operating conditions. Relative measurement errors of all parameters are statistically summarized in Table 4. The maximum relative error of all monitored indicators is controlled within 2.08%, satisfying high-precision endurance monitoring requirements for automotive RF devices.

Table 4. Relative Measurement Error Calibration Results of the Online Monitoring System

Monitored Parameter	Average Relative Error	Maximum Relative Error
S21 small-signal gain	0.72%	1.35%
PAE	0.95%	2.08%
Static drain current I_{DQ}	0.41%	0.87%
Junction temperature T_j	0.33%	0.69%
THD total harmonic distortion	1.12%	1.94%

4. Multi-Domain Degradation Feature Extraction and Quantitative Characterization via CPDI Comprehensive Performance Degradation Index

4.1 Time-Series Multi-Domain Degradation Feature Extraction

Raw noise interference is eliminated and weak degradation trends are amplified by extracting degradation features from three dimensions based on long time-series datasets collected by the platform:

1) **Time-domain degradation features:** linear degradation slope k_i, second-order curvature c_i, long-term steady-state drift offset ΔX_i of each parameter over aging time;
2) **Frequency-domain RF degradation features:** full-band S-parameter offset integral, cumulative harmonic power increment, average deterioration amplitude of IMD3;
3) **Time-frequency joint features:** parameter recovery time constant τ_i after temperature disturbance, used to distinguish reversible drift (τ_i<5 min) and permanent irreversible damage (τ_i>60 min).

4.2 Failure Sensitivity Analysis of Parameters Based on Grey Relational Analysis

The grey relational analysis method is adopted to calculate the correlation degree r_i between each characteristic parameter and device failure severity; higher correlation degrees indicate stronger sensitivity of the parameter to micro-damage variations. Full time-series data of 10 completely failed PAs are used for correlation calculation. Six high-sensitivity indicators with $r_i \geq 0.75$ are retained as input features for CPDI: S21 gain, PAE, I_{DQ}, I_G, THD, and junction temperature rise amplitude (Lee C, Park J & Kim S., 2023).

4.3 Construction of CPDI Comprehensive Performance Degradation Index

4.3.1 Parameter Normalization Processing

Parameters are divided into positive performance indicators (higher values correspond to healthier devices: S21, PAE) and negative degradation indicators (higher values correspond to more severe damage: I_{DQ}, I_G, THD, temperature rise). Differentiated normalization formulas are applied:

For positive performance indicators: $x_i^*=\frac{X_{i,max}-X_i}{X_{i,max}-X_{i,min}}$

For negative degradation indicators: $x_i^*=\frac{X_i-X_{i,min}}{X_{i,max}-X_{i,min}}$

where X_i is the real-time monitoring value; $X_{i,\max}$ and $X_{i,\min}$ denote upper and lower limits permitted by automotive standards. All normalized features $x_i^* \epsilon [0,1]$, with larger values representing more severe degradation of the corresponding indicator.

4.3.2 Weight Assignment Coupling Failure Mechanisms and Grey Correlation Degrees

The weight w_i of each parameter is computed by coupling the damage contribution ratio of failure mechanisms and grey correlation degree r_i, subject to the constraint $\sum_{i=1}^{6} w_i=1$. Calibrated weights of input features are shown in Table 5.

Table 5. Calibrated Weights of CPDI Input Characteristic Parameters

Characteristic Parameter	Grey Correlation Degree r_i	Mechanism Damage Contribution Ratio	Coupled Weight w_i
S21 gain	0.92	0.24	0.23
PAE	0.89	0.22	0.21
Static drain current I_{DQ}	0.84	0.18	0.17
Gate leakage current I_G	0.81	0.16	0.15
THD harmonic distortion	0.78	0.12	0.11
Junction temperature rise ΔT_j	0.76	0.08	0.13

4.3.3 Complete Calculation Formula of CPDI

The comprehensive performance degradation index is derived by integrating normalized features and coupled weights:

CPDI ranges within [0,1]. Values approaching 0 indicate undamaged devices in healthy operating conditions, while values approaching 1 represent complete endurance failure. Periodic minor fluctuations of the CPDI time-series curve only reflect reversible thermal drift, whereas continuous monotonic growth signifies cumulative irreversible micro-damage.

5. FP-LSTM Degradation Prediction Model Constrained by Failure Physics

5.1 Limitations of Conventional Degradation Prediction Models

1) **Pure physics-of-failure models:** Only applicable to single-stress operating conditions, incapable of characterizing nonlinear coupled degradation of multi-parameters under composite temperature-humidity-power stresses, leading to high fitting errors.
2) **Pure LSTM data-driven models:** Absent of device physical mechanism constraints, susceptible to overfitting with small-scale accelerated aging datasets and lacking generalization across new PA models.

To address the above defects, this paper develops an FP-LSTM hybrid model consisting of a bottom failure physics baseline constraint layer and an upper LSTM time-series residual fitting layer.

5.2 Overall Architecture of the FP-LSTM Hybrid Model

Two layer coupling structure: bottom layer failure physical mechanism layer, upper layer temporal LSTM data fitting layer;

1) **Failure physical layer:** Construct a composite multi stress coupled degradation benchmark model, output the theoretical degradation trend as the prior constraint term of LSTM; Composite stress degradation rate model: coupling temperature, humidity, RF power, and vibration four stress acceleration factors;
2) **LSTM temporal learning layer:** Input online monitoring temporal CPDI sequence, learn the residual nonlinear law of actual degradation and theoretical physical model;
3) **Output layer:** Predict the CPDI attenuation value and remaining durability life RUL for any future aging duration.

5.3 Dataset Partitioning and Hyperparameter Optimization

Full time-series CPDI datasets collected from 24 PAs under 1000-hour aging are split into training, validation, and test sets at a 7:2:1 ratio. Grid search is implemented for LSTM hyperparameter optimization. The optimal combination is determined as: 64 LSTM units, 120 training epochs, initial learning rate of 0.001, and physical constraint penalty coefficient of 0.15. A failure physics deviation penalty term is added to the loss function to suppress model outputs violating device degradation physical laws. (Smith A, Johnson B & Williams D., 2021).

5.4 Comparative Evaluation Metrics for Benchmark Models

Four conventional models are selected for comparison: Model 1 – single-stress Arrhenius physics-of-failure model; Model 2 – unconstrained pure LSTM; Model 3 – BP neural network; Model 4 – linear degradation fitting model. Three quantitative metrics are adopted to evaluate prediction accuracy:

1) Mean Absolute Error (MAE): $MAE=\frac{1}{N}\sum_{i=1}^{N} \mid \hat{CPDI_i}-CPDI_i \mid$
2) Root Mean Square Error (RMSE): $RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N} (\hat{CPDI_i}-CPDI_i)^2}$
3) Mean Relative Error (MRE): $MRE=\frac{1}{N}\sum_{i=1}^{N} \frac{\mid \hat{CPDI_i}-CPDI_i \mid}{CPDI_i}\times 100\%$

Quantitative prediction performance of all models on the test set is summarized in Table 6.

Table 6. Prediction Accuracy Comparison of Different Degradation Models

Prediction Model	MAE	RMSE	MRE / %
Single-stress Arrhenius Physics-of-Failure Model	0.0742	0.0915	60.69
Unconstrained Pure LSTM	0.0146	0.0183	5.82

Prediction Model	MAE	RMSE	MRE / %
BP Neural Network	0.0213	0.0267	8.41
Linear Fitting Model	0.0581	0.0724	42.37
Proposed FP-LSTM Hybrid Model	0.0079	0.0098	3.17

The proposed FP-LSTM model achieves optimal performance across all three error metrics. The mean relative error is reduced by 47.52% compared with the traditional single-stress Arrhenius model. Physical prior constraints effectively mitigate overfitting of pure LSTM networks and significantly improve degradation prediction accuracy under coupled multi-stress conditions.

6. Four-Tier Hierarchical Comprehensive Endurance Degradation Evaluation Method for Automotive-Grade RF PAs

6.1 Calibration Basis of the Evaluation System

The evaluation framework is established based on AEC-Q102 RF device endurance standards, ISO 26262 automotive functional safety failure thresholds, and calibrated CPDI time-series curves from 10 fully aged Pas (Wang J, Zhang Q & Li H., 2022). Core judgment criteria include real-time CPDI degradation index and time-frequency joint feature recovery time constant τ_i, which automatically distinguish reversible thermal drift from irreversible micro-damage and eliminate misjudgment induced by temperature fluctuations.

6.2 Quantitative Four-Tier Endurance State Classification Thresholds

Three critical boundary thresholds T_1=0.22 , T_2=0.56, T_3=0.84 are calibrated using full aging time-series data of 24 PA samples, dividing device endurance performance into four tiers. Classification criteria are presented in Table 7.

Table 7. Four-Tier Hierarchical Evaluation Criteria for PA Endurance Performance

Endurance Tier	CPDI Range	Device Damage State	Engineering Application Recommendation
Tier 1: Healthy State	$0 \leq CPDI \leq 0.22$	Only reversible temperature drift exists; no permanent micro-interface damage	Satisfies 15-year full-vehicle endurance service requirements
Tier 2: Slight Irreversible Degradation	$0.22 < CPDI \leq 0.56$	Minor accumulation of interface states, slight attenuation of RF performance without safety hazards	Eligible for short-term service; early warning triggered on production lines
Tier 3: Critical Degradation State	$0.56 < CPDI \leq 0.84$	Progressive electromigration and gate corrosion, severe deterioration of linearity and efficiency	Prohibited for long-term vehicle integration; premature replacement required
Tier 4: Complete Endurance Failure	$CPDI > 0.84$	Micro-damage accumulates to critical thresholds; RF indicators exceed functional safety limits	Confirmed endurance failure; unqualified for vehicle installation

6.3 Standardized Online Endurance Evaluation Full Workflow

1) Continuous loading of coupled multi-stress endurance aging; millisecond-level synchronous collection of full-dimensional degradation parameters via the monitoring platform.
2) Preprocessing of raw time-series data: noise filtering, RF link error compensation, extraction of reversible drift features, and

calculation of temperature disturbance recovery time constant τ_i.

3) Reversible thermal drift interference filtering: If τ_i<5 min for all parameters, the device is judged free of permanent damage and directly classified as Tier 1 healthy state.
4) Extraction of six high-sensitivity features, normalization, and weighted calculation of real-time CPDI comprehensive degradation index.
5) Real-time prediction of CPDI evolution trends and Remaining Useful Life (RUL) at 300 h, 600 h, and 1000 h in the future via the FP-LSTM model.
6) Classification of current device endurance tier against four-tier threshold standards.
7) Reverse tracing of dominant failure mechanisms based on degradation contribution weights of each characteristic parameter, and automatic generation of standardized AEC-Q102 endurance aging test reports.

7. Comparative Accelerated Aging Experiments and Result Analysis

7.1 Test Samples and Grouping Scheme

A total of 24 commercial automotive PA devices are adopted, including 12 GaAs pHEMT 24 GHz radar PAs and 12 GaN HEMT high-power PAs for 77 GHz radar, with output power ranging from 18 to 32 dBm. Three parallel test groups are set up:

- **Control Group A:** Conventional offline segmented aging scheme; power supply cut off and devices cooled every 50 h for discrete single-point testing via offline VNA.
- **Control Group B:** Simplified single-DC online monitoring scheme; only I_{DQ} collected without multi-dimensional synchronous RF acquisition.
- **Test Group C:** The proposed integrated framework combining full online monitoring, CPDI quantitative evaluation, and FP-LSTM prediction.

Unified aging stress conditions: −40 ℃ to 125 ℃ temperature cycling + 85 ℃/85% RH temperature-humidity bias + continuous 77 GHz high-power RF excitation + triaxial random vibration. Total aging duration reaches 1000 hours, with a full-dimensional monitoring dataset stored every 10 minutes.

7.2 Comparative Analysis of Time-Series Data Integrity

Only 20 discrete time-node performance datasets are obtained via offline testing for Control Group A, leaving extensive blank intervals on degradation curves. Three samples exhibit abrupt CPDI growth inflection points at 220 h (Brown R & Miller T., 2022), 470 h, and 730 h corresponding to accelerated micro-damage evolution, which are completely missed by the offline scheme. Test Group C fully records 14400 groups of continuous time-series data, enabling precise localization of all abrupt degradation inflection points. Early latent failure detection is advanced by 118–134 hours compared with offline testing, effectively preventing omission of latent damage.

7.3 Validity Verification of the CPDI Comprehensive Degradation Index

A 77 GHz GaN automotive PA is subjected to a 1000-hour full time-series aging test to track the correlation between CPDI evolution and endurance state. Key experimental observations are listed as follows:

- **0–240 h**: CPDI stabilizes within 0.15–0.21, with only periodic minor fluctuations induced by ambient temperature variations and no persistent upward trend; the device is classified as Tier 1 healthy.
- **240–510 h:** CPDI rises monotonically and remains within 0.23–0.54, indicating slight irreversible degradation corresponding to Tier 2.
- **510–820 h:** CPDI climbs to 0.57–0.82 with aggravated performance degradation, entering Tier 3 critical degradation state.
- **Post 820 h:** CPDI exceeds the 0.84 failure threshold, with fully accumulated micro-damage confirming Tier 4 endurance failure.

By contrast, under the conventional single-index evaluation method, the S21 gain of the tested sample only decreases by 0.92 dB at the 680-hour aging node, failing to reach the industry-standard 1 dB failure threshold and resulting in false healthy judgment. The comparative experiment validates the superiority of multi-dimensional fused CPDI, which reliably distinguishes reversible thermal drift from permanent micro-damage and achieves a 98.6%

overall recognition accuracy for irreversible device degradation, far outperforming single-indicator evaluation schemes.

7.4 Engineering Validation of FP-LSTM Lifespan Prediction

Six PA samples from the test set are selected for remaining lifespan prediction at early aging nodes of 200 h, 400 h, and 600 h. The proposed FP-LSTM model delivers a maximum prediction deviation of 32 hours for remaining useful life, while the traditional Arrhenius model exhibits a maximum early-stage prediction deviation of 168 hours. The results verify the superior reliability of the physics-constrained time-series network under small-sample early aging conditions, supporting advance prediction of device endurance lifespan on production lines and optimizing automotive device screening workflows.

7.5 Case Validation of the Four-Tier Hierarchical Evaluation System

Three typical failure samples are selected for mechanism tracing verification:

1) **Electromigration-dominated failure:** I_{DQ} and PAE contribute over 60% to total CPDI degradation weight; the evaluation system automatically traces metal interconnection degradation as the root cause.
2) **Gate corrosion under humid conditions:** I_G and THD occupy the highest degradation contribution weights, identifying gate electrochemical corrosion as the dominant failure mechanism.
3) **Pure reversible thermal drift:** Recovery time constant τ_i=2.1 min, directly classified as Tier 1 healthy without permanent damage.

Tier classification and failure mechanism tracing results for all three samples perfectly match post-test Scanning Electron Microscopy (SEM) micro-morphology observations of dissected devices, demonstrating robust engineering reliability of the complete evaluation framework.

8. Conclusions and Future Outlook

8.1 Core Quantitative Conclusions of This Work

1) A multi-stress coupled online synchronous monitoring platform for automotive RF PA endurance aging is constructed with a nanosecond-level time-sharing RF switching timing architecture. Multi-parameter synchronous acquisition of DC metrics, RF characteristics, distortion, and junction temperature is realized without interrupting high-power aging stresses. The maximum relative measurement error of all parameters is controlled within 2.08%. The platform fully captures early latent degradation inflection points missed by offline testing, advancing latent failure detection by an average of 126 hours.
2) The CPDI comprehensive performance degradation index is established via coupled weighting of grey correlation degrees and failure mechanisms, enabling quantitative degradation characterization normalized within the range of 0–1. The index automatically distinguishes reversible thermal drift from permanent micro-damage and achieves a 98.6% recognition accuracy for irreversible device degradation, resolving the risk of missed failure judgment induced by single RF indicator evaluation.
3) The FP-LSTM hybrid degradation prediction model embeds composite-stress failure physics models as prior constraints into LSTM time-series networks. Under composite accelerated aging conditions, the model delivers a mean relative error of 3.17% for lifespan prediction, representing a 47.52% accuracy improvement over the traditional single-stress Arrhenius model. Generalization performance under small-sample early aging scenarios is significantly enhanced.
4) Standardized four-tier endurance degradation thresholds are defined based on extensive accelerated aging calibration. A complete standardized workflow covering data acquisition, preprocessing, quantitative characterization, lifespan prediction, tiered judgment, and failure mechanism tracing is established, fully compliant with AEC-Q102 and ISO 26262 automotive device reliability qualification standards, and readily integrable into semiconductor production line aging test equipment.

8.2 Future Research Directions

1) Expand online monitoring methodologies for multi-channel System-in-Package (SiP) integrated RF front-end modules, incorporating channel electromagnetic coupling degradation correction models to adapt to endurance testing of integrated vehicle-mounted RF transceivers.

2) Optimize the FP-LSTM training pipeline via transfer learning: pre-train the model using aging datasets of mature PAs to drastically reduce sample quantities required for calibration of new PA models and shorten production line qualification cycles.
3) Integrate in-situ infrared thermal imaging and offline SEM micro-morphology observation data to construct a multi-scale coupled failure physics model linking micro-morphology and macroscopic electrical performance, further improving degradation prediction and failure mechanism tracing precision.
4) Conduct real-vehicle road endurance online monitoring experiments to establish a lifespan mapping model between laboratory accelerated aging and natural full-vehicle service aging, enabling direct conversion of laboratory test results to equivalent 15-year full-vehicle endurance performance assessments.

References

Brown R, Miller T. (2022). Limitations of single gain degradation failure criteria for automotive power amplifier qualification // 2022 European Microwave Conference (EuMC). *Milan: IEEE*, 567-570.

Lee C, Park J, Kim S. (2023). Continuous S-parameter online monitoring system for automotive PA accelerated aging test // 2023 IEEE International Reliability Physics Symposium (IRPS). *Monterey: IEEE*, 4A.3-4A.7.

Smith A, Johnson B, Williams D. (2021). Reliability characterization of GaN HEMT power amplifiers for 77 GHz automotive radar applications. *IEEE Transactions on Electron Devices, 68*(5), 2112-2119.

Wang J, Zhang Q, Li H. (2022). Offline accelerated aging test methodology compliant with AEC-Q102 for automotive RF power transistors. *Microelectronics Reliability, 126*, 114356.

Zhang H, Liu Y, Wang L. (2023). Remaining useful life prediction of RF power amplifiers based on optimized LSTM network under multi-stress aging conditions. *Reliability Engineering & System Safety, 239*, 109876.