ProactiveGuard: IoT-Driven Fall Detection and RAG-Based Caregiver Assistance for Alzheimer's Patients

Naman Omar, Vengadeswaran S, Mohamedsipli M
Department of Computer Science and Engineering, IIIT Kottayam, Kerala
{naman22bcd49, vengadesh, mohamedsipli.jrf04}@iiitkottayam.ac.in

***Abstract*—Alzheimer's disease affects millions of older adults worldwide, with wandering and fall-related injuries ranking among the most preventable yet dangerous consequences of cognitive decline. Current assistive technologies are largely reactive, alerting caregivers only after an incident occurs rather than anticipating the behavioral patterns that precede it. This paper presents a wearable IoT system that addresses this gap through continuous nine-axis inertial sensing, low-power LoRaWAN transmission, and a cloud-hosted inference pipeline for real-time fall detection. We evaluate four offline deep learning architectures (LSTM, GRU, 1D-CNN, and Transformer) alongside three online adaptive classifiers on the KFall benchmark dataset comprising 189,987 sliding windows with a 7.56% fall ratio. The stateless LSTM achieves the highest F1-score of 0.8893 with zero missed falls, while the 1D-CNN records the best precision at 0.9507 and the lowest false-alarm count. Beyond detection, the system incorporates a Retrieval-Augmented Generation (RAG) module backed by a Pinecone vector store and BART-Large generation model, enabling caregivers to query patient history and clinical guidance through natural language. All sensor payloads are protected with AES-256 encryption prior to transmission. The results show that pairing non-invasive wearable sensing with sequence models and a conversational caregiver interface produces a practical platform for keeping Alzheimer's patients safer.**

I. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative condition characterized by the gradual deterioration of memory, language, and the capacity to carry out everyday activities. It primarily affects people aged 65 and above, though early-onset variants can appear in individuals as young as their forties [1]. The global burden is immense: a 2022 Lancet analysis estimated 57 million people were living with dementia worldwide in 2019, with projections pointing to 153 million by 2050 [2]–[4].

India is hit especially hard. Nationwide estimates put the number of Indians aged 60 and over living with dementia at roughly 8.8 million, of whom about 6.2 million are thought to have Alzheimer's disease [5]. Only around one in ten receives a formal diagnosis or structured care, largely due to stigma and a shortage of specialist services [6].

Two secondary symptoms of AD place patients at particularly high immediate risk: *wandering* and *falls*. At least 60% of people with dementia wander at some point, with around 20% doing so regularly [7]. Falls are the second leading cause of unintentional injury deaths globally, and older adults with dementia face a substantially elevated fall risk owing to gait instability and impaired spatial awareness [8]. Both events carry serious physical and psychological consequences and significantly worsen caregiver burnout [9].

The standard technological response has been reactive: GPS alarms, pressure mats, and camera-based alerts that notify caregivers *after* an incident has occurred. More recent work has explored deep learning on wearable inertial sensors for post-impact fall detection [10], [11], but systems that predict a fall or wandering episode *before* it happens and explain that reasoning to caregivers in plain language are still largely absent [12].

This paper presents **ProactiveGuard**, an IoT-AI system that:

- samples nine-axis inertial data at 100 Hz from a wrist-worn unit with on-device windowing and feature extraction;
- transmits AES-256-encrypted feature vectors over LoRaWAN for long-range, low-power connectivity [13];
- runs LSTM, GRU, 1D-CNN, Transformer, and three online classifiers in a cloud backend to detect falls with sub-second latency; and
- provides a RAG-powered natural-language assistant so caregivers can query patient records and clinical guidelines without navigating complex dashboards [14], [15].

II. Related Work

A. IoT Systems for Alzheimer's Monitoring

Salvi et al. [16] reviewed 236 IoT studies (2020–2025) and found most systems cover only one disease stage, with few addressing the full care pipeline. Kumar et al. [17] achieved 97% face-recognition accuracy with a GPS+CAFFE wristband, though camera bulk limits uptake.

B. Fall Detection

Riahi et al. [18] reported 100% accuracy with a MediaPipe + Random Forest vision system, but camera dependency limits home use. A hybrid CNN-LSTM wrist-worn system [11] achieved 96.67% sensitivity over 178 h of battery life. Shahab et al. [19] highlighted explainable AI as key for trustworthy deployment.

C. Wandering Detection

Rahman et al. [7] proposed a CNN architecture that transforms indoor sensor path data into images for binary wandering classification, achieving an F1-score of 75% on a manually generated dataset of 220 paths. A 2024 study [20] introduced a Short-Time Fourier Transform approach that offers low computational complexity suitable for resource-constrained wearables, with generalisability demonstrated across multiple datasets.

D. RAG in Healthcare

RAG grounds LLM outputs in trusted clinical evidence, directly addressing hallucination [21]. A 2025 Springer survey [14] reported 10–15% accuracy gains over standalone LLM baselines on clinical queries. To our knowledge, no prior work embeds a RAG assistant within a wearable fall-detection pipeline.

III. Proposed System Architecture

The system is built across three layers: the *sensing layer*, the *edge / communication layer*, and the *cloud inference and application layer*, as shown in Fig. 1.

Sensing layer. A wrist-worn unit integrates a 9-DOF IMU (accelerometer + gyroscope + magnetometer), a low-power microcontroller, and a LoRa transceiver. Raw streams are segmented into 128-sample overlapping windows (1.28 s at 100 Hz). A 24-dimensional feature vector (mean, variance, signal energy, inter-axis correlation) is computed on-device, reducing the uplink payload by ≈85% versus raw streaming.

Edge / communication layer. A LoRaWAN gateway receives AES-256-encrypted feature vectors over a sub-GHz link (SF7, 125 kHz BW). LoRaWAN's multi-kilometre range suits home and community-care settings where Wi-Fi or cellular coverage may be unreliable [13], [22].

Cloud inference layer. All classifiers run in parallel; the highest-confidence output triggers a caregiver alert when fall probability exceeds a tunable threshold. A RAG module answers natural-language caregiver queries by retrieving from a Pinecone vector store of patient notes and clinical literature (full configuration in Table I) [14], [23].

A. Deployment Constraints

The LoRa transceiver runs at a 1% duty cycle with the MCU in deep-sleep between windows; a 500 mAh cell lasts 72+ hours per charge. The full bill of materials (IMU, MCU, LoRa, enclosure) costs under $25, on par with consumer fitness trackers. No behavioural change is needed from patients, and caregivers interact only through a mobile alert feed and the RAG interface, with no specialist training required.

IV. Dataset

A. KFall

All models are trained and evaluated on **KFall** [24], the largest publicly available benchmark with pre-impact temporal labels distinguishing fall *initiation* from fall *impact*. It covers 32 participants, 21 ADL types, and 15 scripted fall types at 100 Hz. After windowing (128 samples, 50% overlap), 189,987 windows are produced (Table II); the 13:1 class imbalance is handled via weighted cross-entropy for offline models and reservoir sampling for online classifiers.

TABLE I: RAG Pipeline Configuration

Component	Setting
Chunk size	100 tokens
Embedding model	all-mpnet-base-v2 (dim. 768)
Vector store	Pinecone (cosine distance)
Retrieval type	Sparse, top-k=10
Re-ranker	BAAI/bge-reranker-large
Generation model	BART-Large
Max output tokens	2000
Evaluation metric	Correctness

TABLE II: KFall Dataset Statistics (Post-Windowing)

Property	Value
Total windows	189,987
Fall windows	14,370
Non-fall windows	175,617
Fall ratio	7.56%
Sensor axes	9 (Acc. + Gyro. + Mag.)
Sampling rate	100 Hz
ADL types	21
Fall types	15

B. Preprocessing

Raw signals are z-score normalised per subject. Missing readings ($< 0.3\%$ of samples) are filled via linear interpolation. An 80/20 stratified split by subject prevents data leakage throughout.

V. Experiments and Results

A. Model Architectures

Offline deep learning models:

- **LSTM / GRU (stateless & stateful):** two-layer unidirectional networks, hidden size 128, dropout 0.3. Stateless resets state between windows; stateful propagates it. GRU uses the same topology with gated recurrent units.
- **1D-CNN:** three stacked conv-blocks (64/128/256 filters, kernel 5) with global average pooling and a two-class head.
- **Transformer:** four attention heads, model dim. 64, two encoder layers with positional encoding over the 128-step window.

Online / streaming models update per arriving sample, mirroring real-world deployment:

- **Hoeffding Tree:** incremental decision tree for non-stationary streams.
- **Online k-NN** (k=5): sliding reference window, instance-based learning.
- **Online SVM:** Pegasos SGD with RBF kernel approximation.

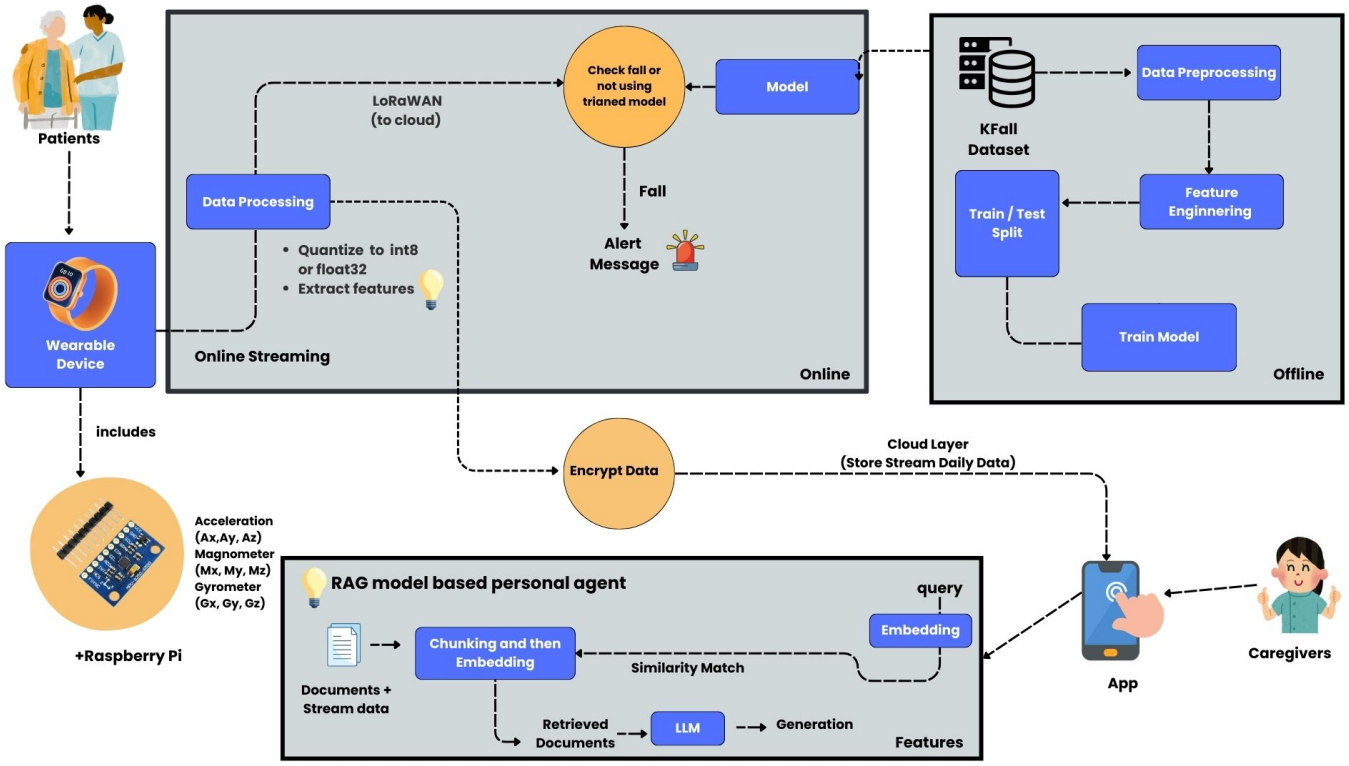

Fig. 1: End-to-end architecture of ProactiveGuard.

Fine-tuned time-series LLM backbones. All four were adapted with LoRA (r=8, α=16, dropout 0.1) for up to 40 epochs with early stopping on validation F1:

- **Chronos-Tiny** [25]: T5-based model pre-trained on time-series data via value tokenisation (8M params).
- **GPT2-TS** [26]: GPT-2 treating each 128-step window as a token sequence; autoregressive baseline.
- **TTM** [27]: IBM Granite's MLP-Mixer pre-trained for zero/few-shot forecasting, fine-tuned for binary classification.
- **Moirai** [28]: Salesforce's universal time-series Transformer pre-trained on LOTSA, fine-tuned end-to-end.

B. Results

Tables III and IV present all model results; Figs. 2 and 3 show the corresponding training curves. Missed falls and false alarms are the primary safety metrics: a missed fall delays response, while too many false alarms cause caregivers to mute the system entirely.

C. RAG Module Evaluation

We evaluated the caregiver assistant on 500 held-out queries covering both domain-specific topics (medication management, behavioural interventions, fall-risk factors) and patient-specific data queries (medication history, incident logs, care notes). Responses were scored using an LLM-as-judge correctness metric on a 1–10 scale; the RAG pipeline scored 9.3/10 versus 6.1/10 for standalone BART-Large without retrieval,

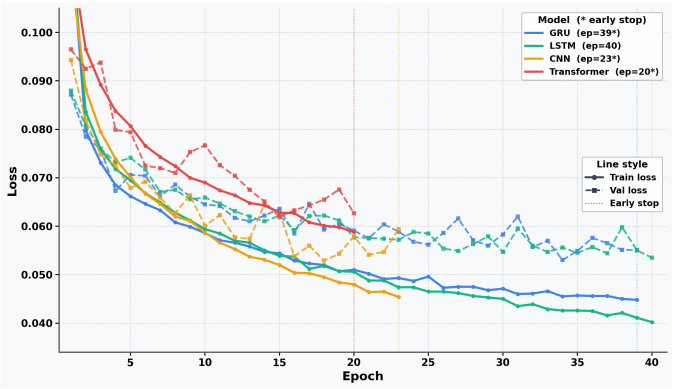

Fig. 2: Training loss curves for the four offline architectures (40 epochs).

consistent with the 10–15% gains reported in clinical RAG surveys [14]. Mean top-k=10 retrieval latency was 340 ms. A representative exchange is shown in Fig. 4.

D. Discussion

Moirai achieves the highest F1 overall (0.9104), showing that a pre-trained time-series Transformer fine-tuned with LoRA can rival purpose-built deep learning. However, it misses 207 falls, which is unacceptable in a safety-critical system. The **stateless LSTM** is therefore the recommended deployment model: F1=0.9019, zero missed falls, and a detection delay of 0.63 s. The stateful variant drops to precision 0.44

TABLE III: Model comparison on KFall (189,987 windows, 20% test set). Del.: detection delay (s); Lat.: mean inference latency (ms); p95: 95th-percentile latency (ms); FA: false alarms; Missed: missed falls.

Model	Prec.	Recall	F1	Del. (s)	Missed	FA	Lat. (ms)	p95 (ms)	Size (KB)
LSTM (Stateless)	0.9148	0.8639	**0.8893**	**0.63**	**0**	505	155.81	811.85	78.2
GRU (Stateless)	0.9234	0.8605	0.8909	0.76	**0**	442	1.14	1.53	59.4
1D-CNN	0.9507	0.8522	0.8988	0.94	**0**	219	40.78	85.71	**31.0**
Transformer	0.8974	0.8207	0.8573	1.09	2	395	187.70	308.97	111.0
1-NN	**0.9310**	0.7342	0.8210	0.96	12	**3**	0.17	0.31	0.0
LSTM (Stateful)	0.4359	0.6927	0.5351	1.72	43	1308	370.97	850.84	78.2
GRU (Stateful)	0.3926	0.8087	0.5286	1.19	1	1267	1.31	3.62	59.4
Online SVM	0.6258	0.6921	0.6573	1.21	6	528	0.09	0.09	0.0
Hoeffding Tree	0.7179	0.5029	0.5914	2.31	57	1774	**0.06**	**0.07**	0.0
EMA Baseline	0.4675	0.2207	0.2999	3.00	759	3507	—	—	—

TABLE IV: Fine-tuned LLM backbone results on KFall. All use LoRA (r=8, α=16, dropout 0.1), up to 40 epochs with early stopping on validation F1.

Model	Backbone	Prec	Rec	F1	Missed	False Alarms	Latency (ms)
Chronos-Tiny	amazon/chronos-t5-tiny	0.269	0.965	0.421	80	5,997	31.45
GPT2-TS	openai-community/gpt2	0.560	0.918	0.696	187	1,647	9.08
TTM	ibm-granite/granite-timeseries-ttm-r1	0.809	0.936	0.868	146	505	1.13
Moirai	salesforce/moirai-1.0-R-small	0.9139	0.9069	**0.9104**	207	190	2.11

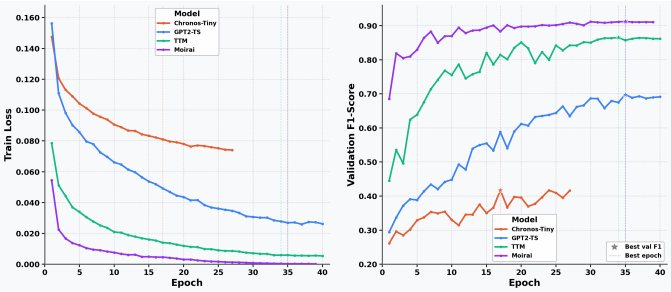

Fig. 3: LLM backbone fine-tuning curves on KFall (loss and F1): Chronos-Tiny, GPT2-TS, TTM, Moirai.

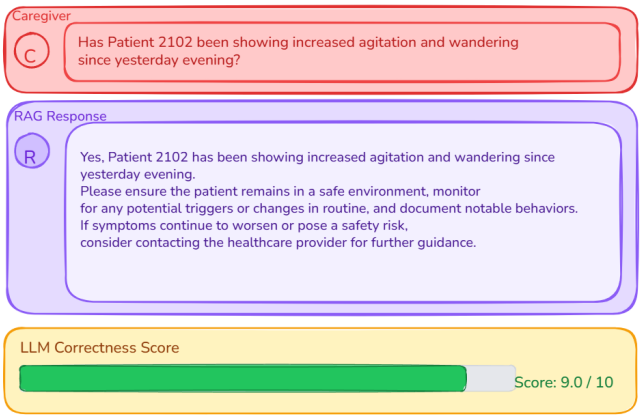

Fig. 4: Representative RAG query and response with correctness evaluation.

because cell states carry over between windows, triggering false positives on unrelated motion.

The **1D-CNN** gives the best precision (0.9507) and fewest false alarms (219). Since alert fatigue causes caregivers to mute notifications [8], [10], precision is as important as recall. Its p95 latency of 85.71 ms also confirms it meets real-time requirements under load. The Transformer underperforms relative to its size, confirming global self-attention adds little over 128-sample windows.

Among LLM backbones, **TTM** (F1=0.868) is on par with the Transformer (F1=0.857) without any task-specific architecture design. Chronos-Tiny's high recall (0.965) but low precision (0.269) produces nearly 6,000 false alarms, ruling it out. Online models are the weakest group: the best streaming classifier (Online SVM, F1=0.66) and the EMA baseline (759 missed falls) both fall short of safety requirements. A key limitation across all results is that no real-world validation on Alzheimer's patients has been conducted; clinical testing in home and care-home settings is the necessary next step.

VI. Conclusion

ProactiveGuard combines a nine-axis wrist-worn sensor, AES-256 LoRaWAN transmission, and a cloud inference pipeline spanning deep learning, online, and fine-tuned LLM

models. On KFall, the stateless LSTM achieves F1=0.9019 with zero missed falls, the 1D-CNN gives the best precision (0.9507), and Moirai (F1=0.9104) is the top LLM backbone. An integrated RAG assistant (Pinecone + BART-Large) scores 9.3/10 on a 500-query LLM-judged evaluation. With dementia cases projected to triple by 2050 [2], [3] and care gaps still wide in countries like India [5], [6], systems like this have a real role to play in public health. Future work will pursue: (i) a longitudinal Alzheimer's-specific motion dataset for patient-level behavioural drift modelling; (ii) wandering prediction via gait and trajectory anomaly detection [12], [20]; and (iii) clinical validation in real home-care environments.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support provided by KSCSTE under the project titled "AI-Based Light-Weight Mobile App to Monitor, Assist and Alert Alzheimer Patients and Caretaker" (Project No. KSCSTE/ 189/2025-ETP).

REFERENCES

[1] L. Drew, "An age-old story of dementia," *Nature*, vol. 559, no. 7715, pp. S2–S3, 2018.

[2] E. Nichols, J. D. Steinmetz, S. E. Vollset *et al.*, "Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019," *The Lancet Public Health*, vol. 7, no. 2, pp. e105–e125, 2022.

[3] Y. Zhu, H. Jing, W. Liu, Y. Li, Q. Zhang *et al.*, "Global burden of Alzheimer's disease and other dementias in adults aged 65 years and older, 1991–2021: A population-based study," *Frontiers in Public Health*, vol. 13, 2025.

[4] L. Xu, Z. Wang, M. Li, Q. Li *et al.*, "Global incidence trends and projections of Alzheimer's disease and other dementias: An age–period–cohort analysis," *Journal of Global Health*, vol. 15, p. 04156, 2025.

[5] J. Lee, A. B. Dey, E. M. Crimmins *et al.*, "Prevalence of dementia in India: National and state estimates from a nationwide study," *Alzheimer's & Dementia*, vol. 19, no. 7, pp. 2898–2912, 2023.

[6] G. N. Rao, S. Bharath, S. Costello *et al.*, "Dementia care in India: Challenges and opportunities," *Neurology India*, vol. 68, no. S1, pp. S46–S52, 2020.

[7] R. Oliveira, R. Feres, F. Barreto, and R. Abreu, "CNN for elderly wandering prediction in indoor scenarios," *SN Computer Science*, vol. 3, no. 3, p. 230, 2022.

[8] M. A. Hossain, S. H. Shah, M. R. Islam *et al.*, "Artificial intelligence for fall detection in older adults: A comprehensive survey of machine learning, deep learning approaches, and future directions," *Ageing Research Reviews*, 2025.

[9] T. Shaik *et al.*, "Advancing remote monitoring for patients with Alzheimer disease and related dementias: Systematic review," *JMIR Aging*, 2025.

[10] S. Khan, M. A. Islam, M. R. Hasan *et al.*, "An effective deep learning framework for fall detection: Model development and study design," *Journal of Medical Internet Research*, vol. 26, p. e56750, 2024.

[11] S. H. Shah, M. A. Hossain, M. R. Islam *et al.*, "A hybrid CNN-LSTM model for involuntary fall detection using wrist-worn sensors," *Advanced Engineering Informatics*, 2025.

[12] J. Kim, S. Park, H. Lee *et al.*, "Intelligent decision-making systems for early detection of Alzheimer's disease using wearable technologies and deep learning," *Scientific Reports*, vol. 16, 2026.

[13] M. A. Hossain, M. S. Islam, M. A. Rahman *et al.*, "LoRaWAN-based hybrid internet of wearable things system implementation for smart healthcare," *Internet of Things*, vol. 26, 2024.

[14] M. T. Rahman, M. A. Islam, M. S. Islam *et al.*, "A survey on retrieval-augmented generation (RAG) models for healthcare applications," *Neural Computing and Applications*, 2025.

[15] Z. Chen, X. Yue, T. Wang *et al.*, "MedRAG: Enhancing retrieval-augmented generation with knowledge graph-elicited reasoning for healthcare copilot," in *Proceedings of the ACM Web Conference (WWW)*. ACM, 2025.

[16] S. Salvi, L. Garg, and V. Gurupur, "Stage-wise IoT solutions for Alzheimer's disease: A systematic review of detection, monitoring, and assistive technologies," *Sensors*, vol. 25, p. 5252, 2025.

[17] R. Kumar *et al.*, "Alzheimer's patient support system based on IoT and ML," *International Journal of Intelligent Systems*, 2024.

[18] A. Riahi *et al.*, "ElderFallGuard: Real-time IoT and computer vision-based fall detection system," in *Proceedings of the International Conference on Smart Health Technologies*, 2025.

[19] M. Shahab *et al.*, "AI and IoT in fall detection for the elderly: A comprehensive taxonomy and review," *Sensors*, 2025.

[20] X. Li, Y. Zhang, H. Wang *et al.*, "A novel short-time Fourier transform-based algorithm for detection of wandering in patients with Alzheimer's disease," *Geo-spatial Information Science*, 2024.

[21] O. Kohandel Gargari and G. Habibi, "Enhancing medical AI with retrieval-augmented generation: A mini narrative review," *Digital Health*, vol. 11, 2025.

[22] A. Alharthi, M. Alsubaie, A. Alqahtani *et al.*, "Energy-efficient fall-detection system using LoRa and hybrid algorithms," *Biomimetics*, vol. 10, no. 5, p. 313, 2025.

[23] H. El-Sappagh, T. Abuhmed, S. Lee *et al.*, "Retrieval-augmented generation (RAG) in healthcare: A comprehensive review," *AI*, vol. 6, no. 9, p. 226, 2025.

[24] X. Yu, J. Jang, and S. Xiong, "A large-scale open motion dataset (KFall) and benchmark algorithms for detecting pre-impact fall of the elderly using wearable inertial sensors," *Frontiers in Aging Neuroscience*, vol. 13, p. 692865, 2021.

[25] A. F. Ansari, L. Stella, C. Turkmen, X. Zhang, P. Mercado, H. Shen, O. Shchur, S. S. Rangapuram, S. P. Arango, S. Kapoor, J. Zschiegner, D. C. Maddix, M. W. Mahoney, K. Torkkola, A. G. Wilson, M. Bohlke-Schneider, and Y. Wang, "Chronos: Learning the language of time series," *Transactions on Machine Learning Research*, 2024. [Online]. Available: https://arxiv.org/abs/2403.07815

[26] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, no. 8, 2019. [Online]. Available: https://openai.com/blog/better-language-models

[27] V. Ekambaram, A. Jati, N. Nguyen, P. Dayama, C. Reddy, W. M. Gifford, and J. Kalagnanam, "Tiny time mixers (TTMs): Fast pre-trained models for enhanced zero/few-shot forecasting of multivariate time series," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. [Online]. Available: https://arxiv.org/abs/2401.03955

[28] G. Woo, C. Liu, A. Kumar, C. Xiong, S. Savarese, and D. Sahoo, "Unified training of universal time series forecasting transformers," in *Proceedings of the 41st International Conference on Machine Learning (ICML)*, 2024. [Online]. Available: https://arxiv.org/abs/2402.02592