Demo Abstract: CaringFM: An Interactive In-home Healthcare System Empowered by Large Foundation Models

Haiyang Wu, Kaiwei Liu, Siyang Jiang, Zhihe Zhao, Zhenyu Yan, and Guoliang Xing*
The Chinese University of Hong Kong, Hong Kong SAR

ABSTRACT

The demand for fully on-device health monitoring is huge and urgent. However, deploying Large Foundation Models conventionally relies on cloud-based computing services, which poses privacy concerns. Driven by the belief of delivering personalised healthcare to family members, this study presents the development of an innovative on-device machine learning system, CaringFM. This family caring system utilizes privacy-protecting sensors and an edge-deployed Foundation Model(FM) to offer a convenient and low-cost solution for chronic disease prediction and health condition monitoring at home. In particular, CaringFM provides general health suggestions and personalized medical information while ensuring high privacy by processing and preserving all data locally.

KEYWORDS

Large Foundation Model, Human Activity Recognition

1 INTRODUCTION

In-home healthcare monitoring market is estimated to be $1.7 billion by 2027[7]. Meanwhile, recent advancements in deep learning methodologies have yielded substantial successes in the domain of intelligent healthcare systems [2, 8]. However, most existing healthcare monitoring systems can only process very few sensor data due to limited compute resources[3], the emergence of the large foundation models (FMs) like GPT[1] and Llama[4] has catalyzed a paradigm shift in the field of AI-enabled embedded systems, with a marked impact on the sphere of wireless health monitoring with desirable general knowledge about diverse sensor data. However, most such FM-powered systems are cloud-based services due to the intensive compute requirements. To meet the escalating demand for non-invasive health monitoring, we propose a novel in-home healthcare monitoring system, CaringFM, which couples multiple sensors with a FM-powered edge-cloud cooperation. CaringFM aims to facilitate individuals' comprehension of their health status in a user-friendly and cost-effective manner. Specifically, CaringFM can capture and analyze diverse sensor data and offer precise predictions for chronic conditions. The incorporation of FM significantly augments the system's functionality by providing tailored, expert-level health insights to users. This demonstration abstract delineates the architecture and operational dynamics of the CaringFM system.

2 SYSTEM DESIGN AND IMPLEMENTATION

The CaringFM system is developed to monitor human activities, prioritizing privacy considerations. As depicted in Fig. 1, the system's hardware suite encompasses an NVIDIA Jetson ORIN[5] for edge computing of up to 200 trillion operations per second (TOPS), Vzense DS77[6], a privacy-conscious indirect-Time-of-Flight (iToF) depth sensor for High Dynamic Range (HDR) imaging at 15 fps, a thermal sensor, a microphone, and a speaker. The iToF depth sensor is selected for its ability to capture depth maps without disclosing sensitive bio-metric information, such as facial features[9].

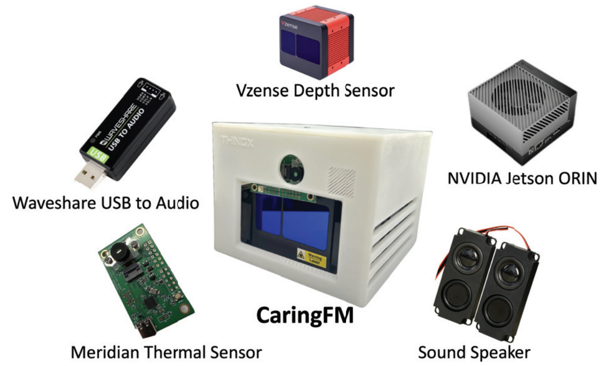

Figure 1: The hardware configuration of CaringFM.

First, depth maps are analyzed by a tiny Human Activity Recognition (HAR) model, which integrates the YOLO[12] and ST-GCN[13] algorithms, to identify actions from 15-frame sequences and supports the recognition of eight distinct activities: Standing, Walking, Sitting, Lying, Standing up, Sitting down, Falling, and Unknown. Second, Incorporating a large foundation model, Rocket-3B, consisting of 3 billion parameters, the CaringFM system functions as an intelligent agent[10]. It accesses an extensive external knowledge database containing detailed health, medical, and user activity data. Finally, utilizing Retrieval-Augmented Generation (RAG) technology, the system enhances its capability to extract relevant information from this database[11]. We combined the medical and activity database into one knowledge-base. Then we embedded the knowledge-base as well as user's utterance into the vector space, computed the cosine similarity between them, and found out top relevant information. The model's parameters, originally in half-precision floating-point (FP16) format, are converted to the bfloat16 format for enhanced computational efficiency on the Jetson ORIN. This conversion reduces the model's memory demands. Consequently, Rocket-3B can produce precise and personalized health information in response to user queries. The system is optimized to enhance key performance indicators, such as recognition accuracy, response latency, and operation efficiency. The experiment result proved our assumptions of deploying a 3B-FM and an HAR model on Jetson ORIN, with recognition accuracy achieved 98%, and the real-time inference rate achieved 10 FPS. We also evaluated the performance of the FM using LLMOps techniques such as Auto-Evaluator from LangChain.

Fig. 2 shows an overall architecture design of CaringFM system.

*Corresponding email: glxing@cuhk.edu.hk.

979-8-3503-6201-5/24/$31.00 ©2024 IEEE
DOI 10.1109/IPSN61024.2024.00028

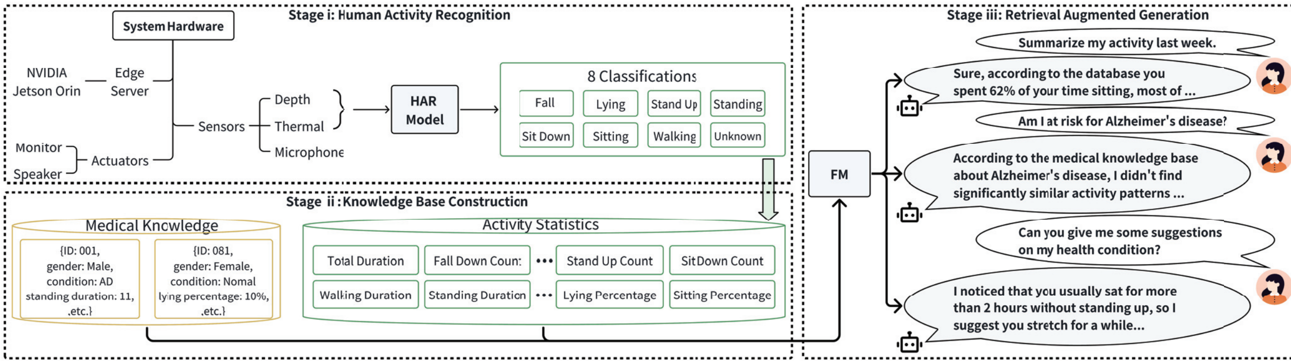

Figure 2: System architecture of CaringFM. In stage i, CaringFM classify the captured sensor data into eight human activities. In stage ii, CaringFM combines activity statistics and medical knowledge to form a knowledge-base. In stage 3, the FM retrieves external information from the knowledge-base and generates responses based on the user's questions.

3 INTERACTIVE DEMONSTRATION

Fig. 3 shows a demonstration of CaringFM with a volunteer falling down. When the subjects move within the Field of View (FoV) of the multi-modal sensors, their activities will be recognised by HAR model, and recorded with a timestamp in the local database. Users can request health information or inquire about their activity data. The voice input captured by the microphone is being processed by a locally deployed speech-to-text model. The system, via the FM, processes these requests and provides text output, which is being processed by a locally deployed text-to-speech model.

During the conference, we will provide an interactive demonstration of the CaringFM system. Participants will have the opportunity to engage with the system and experience its functionalities first-hand. They will be able to communicate with the large foundation model, ask questions, and receive personalised information based on their activity data. The interactive demonstration will showcase a comprehensive understanding of the system's capabilities and its potential to revolutionise family care. A demo video of CaringFM is available at https://shorturl.at/fJVW7.

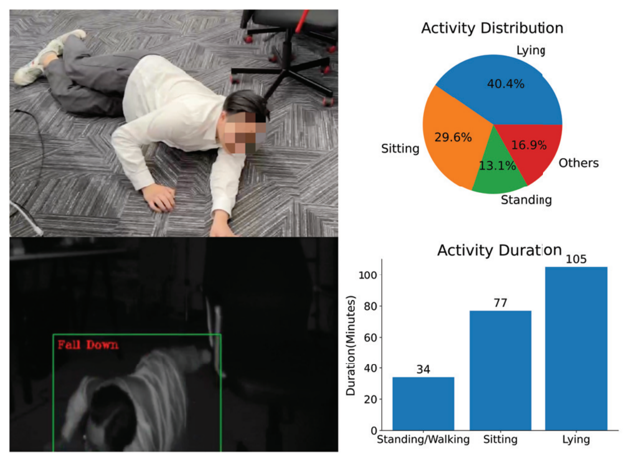

Figure 3: A simulated falling is acted out by a volunteer. The dashboard displays activity statistics retrieved from knowledge-base by FM.

4 CONCLUSION

The CaringFM system, which integrates various privacy-conscious sensors with a large foundation model, affords precise and longitudinal family health monitoring. This demo shows that CaringFM can successfully monitor human activities and provide insightful suggestions through retrieving user's daily activity data as well as the professional medical data.

5 ACKNOWLEDGEMENT

This work is supported by Hong Kong RGC C4034-21G.

REFERENCES

[1] Tom B. Brown, Benjamin Mann, Nick Ryder, and et al. 2020. Language Models are Few-Shot Learners. arXiv:cs.CL/2005.14165
[2] Wei Ding, Siyang Jiang, Hsi-Wen Chen, and Ming-Syan Chen. 2023. Incremental Reinforcement Learning with Dual-Adaptive ε-Greedy Exploration. *AAAI* (2023).
[3] Ali Farman, El-Sappagh Shaker, and et al. 2021. An intelligent healthcare monitoring framework using wearable sensors and social networking data. *Future Generation Computer Systems* 114 (2021), 23–43.
[4] Touvron Hugo, Lavril Thibaut, and et al. 2023. LLaMA: Open and Efficient Foundation Language Models. arXiv:cs.CL/2302.13971
[5] NVIDIA Inc. 2024. NVIDIA Jetson ORIN. https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-orin/.
[6] Vzense Inc. 2024. Vzense DS77 Lite. https://www.vzense.com/cw-itof/ds77.
[7] Insider Intelligence. 2023. The technology, devices, and benefits of remote patient monitoring in the healthcare industry. https://www.insiderintelligence.com/insights/remote-patient-monitoring-industry-explained/.
[8] Siyang Jiang, Rui Fang, Hsi-Wen Chen, Wei Ding, and Ming-Syan Chen. 2023. Dual Adversarial Alignment for Realistic Support-Query Shift Few-shot Learning. *arXiv preprint arXiv:2309.02088* (2023).
[9] Xiaomin Ouyang, Xian Shuai, Li Pan, Xifan Zhang, Heming Fu, Xinyan Wang, Shihua Cao, Jiang Xin, Hazel Mok, Zhenyu Yan, Doris Sau Fung Yu, Timothy Kwok, and Guoliang Xing. 2024. ADMarker:A Multi-Modal Federated Learning System for Monitoring Digital Biomarkers of Alzheimer's Disease. *arXiv preprint arXiv:cs.LG/2310.15301* (2024).
[10] Pansophic. 2024. Rocket-3B. https://huggingface.co/pansophic/rocket-3B.
[11] Lewis Patrick, Perez Ethan, and et al. 2021. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv:cs.CL/2005.11401
[12] Joseph Redmon and Ali Farhadi. 2017. YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.
[13] Sijie Yan, Yuanjun Xiong, and Dahua Lin. 2018. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. arXiv:cs.CV/1801.07455