Vision transformer-based emergency classification in Azure cloud

Lin Cho, Seongho Kim, Eontae Kim, Seonggyeong Kim, Wooyoung Son, and Byungseok Kang\*

*Human IT Education Center  
Seoul, Republic of Korea*{lin87438743, hohoho3077, hft055, tjdrud0224, thsdndud000, anferneekang}@gmail.com

*Abstract*—In modern society, due to the rapid aging of the population, health and long-term care (LTC) management of the elderly aged 65 and over are important social issues. Existing machine learning models that analyze standardized features have limitations in effectively interpreting complex bio-signal patterns. To solve this problem, we propose a vision transformer-based emergency classification service. A new foundation model based on the Transformer architecture can understand human language more accurately and classify requests with much less training than before. We modified the existing Transformer model MobileHART to be suitable for emergency classification.

Keywords—vision transformer, emergency classification, MobileHART, emergency alarm, Azure cloud

# Introduction

Recently, due to the rapid aging of the population in Korea, health issues and LTC management of the elderly have emerged as important social issues [1]. In addition, most elderly people are unable to respond immediately when an emergency occurs. According to public research, many elderly people want to live autonomously without any location constraints, so the use of wearable devices is essential [2]. The currently operating bio-signal based emergency classification system for the elderly utilizes wearable sensors, non-contact radar sensors, and IoT-based monitoring systems. In addition, it detects the real-time health status of the elderly, and IBM Watson Health and Avadin are introducing a high-quality commercialized elderly care system in cooperation with hospitals. Inertial Measurement Unit (IMU)-based bio-signal monitoring system presents a new methodology that utilizes deep learning to detect and analyze respiratory status in real time. This method greatly contributes to the early detection of emergency situations [3].

However, existing geriatric emergency triage systems have several limitations. According to previous studies [4], past machine learning models that rely on standardized feature-based analysis have limitations in effectively interpreting complex bio-signal patterns. Existing methods have limitations, especially in analyzing noisy electrocardiogram (ECG) signals. In addition, [5] pointed out the problem of false alarms that occur in emergency triage systems. First, it is difficult to gain the trust of medical staff due to the lack of explanation, which is important in medical AI. Second, most emergency triage systems are optimized for hospitals or indoor environments. Finally, it is difficult to respond immediately to emergency situations (falls, fainting, cardiac arrest, etc.) that occur outdoors.

To overcome the limitations of the past, the transformer-based architecture has recently attracted attention in the field of artificial intelligence. The IBM Natural Language Understanding (NLU) engine uses a new foundation model based on the transformer architecture to understand human language more accurately. It can effectively classify keywords in paragraphs with much less training than the existing models. The Korea police office developed a Transformer-based false report detection model in an emergency reporting system, which greatly improved the accuracy of emergency response. The post-processing process was carried out together with the police to solve small problems. In addition, according to research in the field of robots [6], the Transformer architecture has been proven to be suitable for real-time collision detection due to its fast inference time and high accuracy.

In this study, the Lightweight MobileHART model, which is an improved version of the transformer-based MobileHART model, was used as an emergency classification model. In order to perform effective analysis in a public cloud environment, the Life Care service was implemented in a Microsoft Azure cloud environment. By collecting and analyzing Human Activity Recognition (HAR) and Personal Health Record (PHR) data in real time, more accurate emergency situation detection is possible than the existing model. The proposed model was introduced to increase the transparency of the decision-making process, and a search engine based on the AI ​​model RAG was utilized. The service we propose includes a function that automatically suggests a response method to an abnormality by referring to the safety guidelines based on bio-signals. This can help the elderly to request help more effectively in emergency situations to ensure their health and safety.

# Proposed Model and Services

## MobileHART

The Transformer model was first introduced in Attention Is All You Need. It can have long-range dependencies unlike CNN through the Attention mechanism, and for this reason, it began to be widely used in fields that use time series input data such as text and audio. ViT [7] was used in the existing NLP and Sound domains, to the Vision domain. ViT was mainly created and developed to perform image recognition, classification, and segmentation. There are two major disadvantages of Vision Transformers: the structure is complex compared to the performance of CNN models mainly used in the image field, so the model is heavy and has a lot of computation, and the amount of pre-training data is very large to achieve performance higher than CNN.

Since then, ViT models have been developed to solve the computational problems that arise when moving to the Vision domain through local self-attention to resolve the shortcomings, and to obtain spatial inductive bias such as T2T-ViT [8] and PVTv1[9]. Although the performance has improved, it requires more parameters, and it still has the limitation of being sensitive to data augmentation. Therefore, CNN and Transformer are used together to reduce the model size and obtain local information and global context together in one model.

MobileHART [10] is a model modified from MobileViT [11] to use various types of sensor data rather than images. MobileViT is a model designed to use a transformer model in a limited computing environment (e.g. edge device), and is a model that achieves better top-1 accuracy with 10-75% fewer parameters than other ViT series models. MobileHART is modified to use time series sensor data frames as tokens instead of image patches, and uses multi-headed attention for each sensor to obtain local self-attention for each sensor.

## TabNet

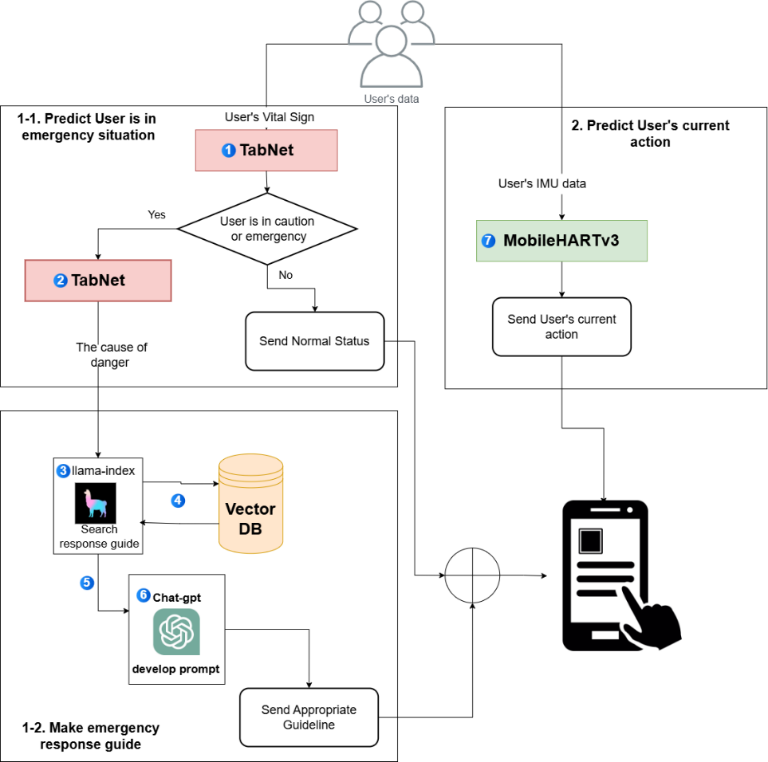
Typically, tree-based techniques (XGBoost [12], LightGBM [13]) and neural network-based models (MLP) [14] are used in tabular data analysis. Tree-based models have shown high performance by utilizing decision-making structures that consider the relationships between features, but they have limited interpretability and have limitations in continuous learning (online learning). On the other hand, neural network-based models (MLP) can be used in high-dimensional data, but there was a problem that they could not effectively learn decision boundaries based on hyperplanes in tabular data.

TabNet is a neural network model dedicated to table data that combines the characteristics of tree models and the advantages of neural network models to solve these problems. It applies the Sparse Feature Selection technique that automatically selects important features for each data sample by utilizing the Sequential Attention Mechanism. In other words, unlike the existing MLP that considers all features equally and learns them, TabNet operates in a way that prioritizes learning meaningful features for each sample. This learning method provides the advantage of increasing the efficiency of the model and enhancing interpretability by excluding unnecessary features.

In this study, we performed bio-signal based emergency situation determination and cause analysis by utilizing TabNet multi-class classification and multi-label classification functions. Afterwards, we combined the search-based RAG (Retrieval-Augmented Generation) model to provide additional information on the cause of the emergency situation, and built a more accurate and interpretable medical response solution.

## Emergency Classification Service

Proposed system architecture supports the entire process from data processing to service deployment based on the Azure cloud platform. Figure 1 is a diagram of the entire system structure and data flow, showing the entire process from data collection to final service deployment.



1. Overall process of emergency classification services.

Our system detects emergency situations in real time based on bio-signal data and provides immediate notifications to users and guardians. It analyzes various bio-data such as the user's heart rate, respiration rate, blood oxygen concentration, skin temperature, and user movement by utilizing wearable sensors and IoT-based data collection devices to classify emergency situations. Unlike existing machine learning-based analysis, the transformer model is applied to enable more precise pattern analysis and abnormality detection. In addition, this service is linked to an emergency reporting system, providing a function that allows users to automatically receive reports when they are in an emergency situation. Guardians can monitor the user's status in real time through a mobile application.

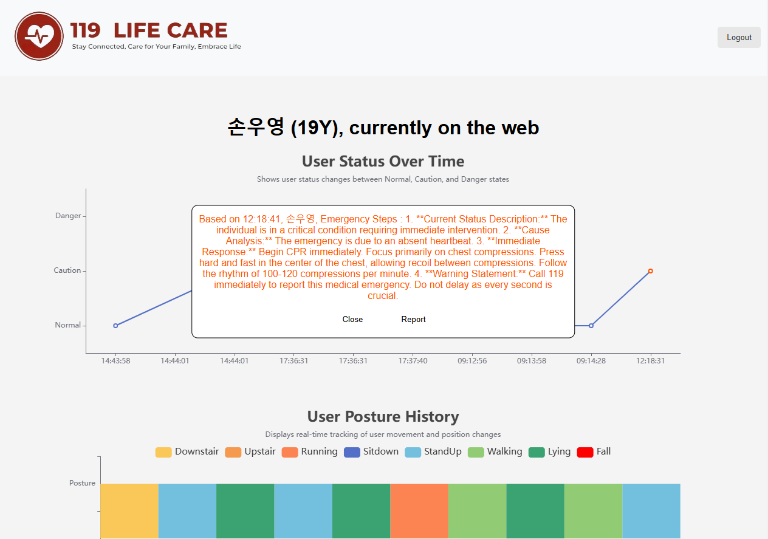
The service largely consists of four stages: data collection, AI analysis, emergency detection and response, and post-monitoring. The detailed process for each stage is as follows. In the first step, the user's biometric data is collected through wearable sensors or IoT-based data collection devices. These devices measure the user's heart rate, breathing rate, blood oxygen concentration (SPO2), etc., and transmit them to the server in real time. This data serves as basic data for comprehensively analyzing the user's condition, and is designed to enable stable data collection even during various activities.

In the second step, the AI ​​model learns the patterns that occur in the data collected from the server, and automatically detects and sends a notification when a specific pattern indicates an emergency. The collected biometric data is analyzed through the Deep Learning-based TabNet model, and one of the user's health conditions (normal, caution, critical) is predicted. Based on the results determined by the AI, the user's risk cause and current condition are synthesized to generate emergency response guidelines based on the WHO.

In the third step, an immediate warning message is sent to the user based on the AI ​​analysis results. If the analysis result is ‘Normal’, monitoring continues without any separate action, but if it is classified as ‘Caution’ or emergency (Danger) status, a notification is sent so that the user can check their own condition and an OpenAI-based emergency response guide is provided. The AI ​​utilizes the OpenAI API to guide the user on how to respond to an emergency. For example, if the blood oxygen level is low, it provides a customized guide in real time according to the user’s condition, such as providing practical response methods such as “open the window immediately and take a deep breath”, so that the user can respond efficiently to an emergency situation.

Finally, continuous monitoring is carried out even after the emergency is over. After a certain period of time after the emergency, the guardian and the user will receive a status check message, and the AI ​​model will analyze the user’s latest status to check if the user’s health condition is stable. It will determine if further action is required. When the emergency is lifted, the service will return to normal mode and continuously monitor biometric data to perform predictive analysis to prevent recurrence. Figure 1 shows our proposed AI model and its service process.

The basic interface displays real-time health monitoring data for logged-in users. The dashboard consists of three major elements: the user information header, the user status chart, and the user posture chart. The user information header displays the user name and age at the top of the screen, and the user status chart is a line chart that tracks the user health status over time, with three statuses: normal (blue), caution (orange), and dangerous (red). If the system detects the user’s posture as “Fall,” an emergency notification pop-up will appear. Red text indicating high severity: "[Current time] [username] has been identified as down. Quick action is required." It appears as a notification message with the phrase: At the bottom of the pop-up notification, select a response with the “Close” button to close the notification and the “Report” button to start filing an emergency report. Figure 2 shows service screen of the fall detection notification.

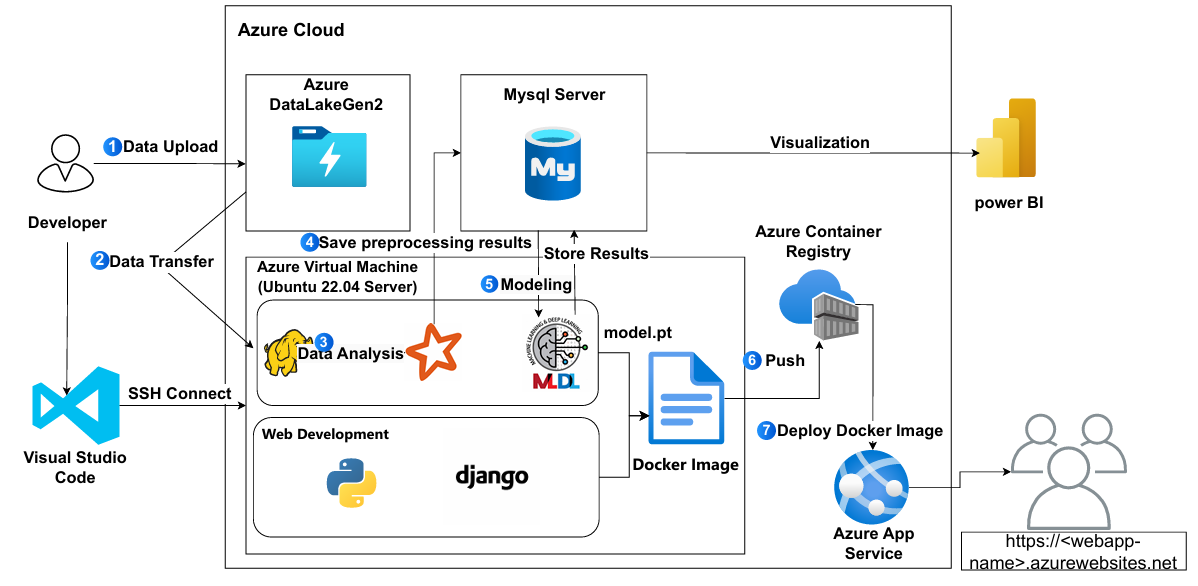


1. Fall detection and user posture history .

# System Architecture and Measurement

## System Architecture

Proposed system architecture supports the entire process from data processing to service deployment based on the Azure cloud platform. Figure 3 is the system architecture for providing bio-signal based emergency classification services. We used Microsoft's Azure cloud to build an effective development environment.



1. System architecture based on Azure cloud.

The core infrastructure of the proposed system is composed as follows. 1) Initial storage and management of large-scale data are performed using Azure Data Lake Gen2. 2) The stored data is transferred to Hadoop Distributed File System (HDFS) built on Azure Virtual Machine, and 3) distributed processing and analysis are performed through Apache Spark. HDFS is a distributed file system of Hadoop for efficient management of large-scale data, and guarantees data stability and processing efficiency by dividing data into blocks and distributing and storing them on multiple nodes.

Apache Spark is an in-memory computing-based data processing engine that overcomes the limitations of the existing Hadoop disk-based processing method and provides fast data processing speed. In the Azure Virtual Machine environment, preprocessing of large-scale data is efficiently performed using the distributed storage system of HDFS and the high-speed processing engine of Spark. 4) The analyzed data is structured and stored in Azure MySQL Server, and 5) Machine/deep learning model is performed based on modeling module.

The bio-signal based emergency classification system application, including the developed MobileHART model and TABNET model, is containerized and registered in 6) Azure Container Registry, and finally distributed through 7) Azure App Service. This cloud-native architecture ensures the scalability and maintainability of the system and enables efficient resource management.

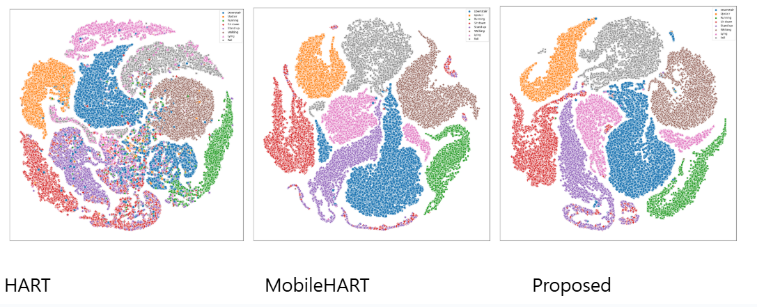
## System Measurements

To evaluate the performance of proposed model, we compared it with two existing HART and MobileHART. We measured five factors that are generally used for evaluation in machine learning domain. Table I shows the results of a comparative analysis of the three models. It can be seen that HART, a simple transformer model, does not learn well and therefore does not have a very high score. The fact that MobileHART has higher accuracy than proposed model seems to be due to overfitting occurring because there is no residual connection. The proposed model outperformed other models in most measurements.

1. Test resuls of three HART base modles

| AI  Models | Measurement Criteria | | | | |
| --- | --- | --- | --- | --- | --- |
| Fall all  [15] | UCI  [16] | HHAR  [17] | PAMAP  [18] | HIFD  [19] |
| HART | 75.43 | 94.42 | 87.76 | 89.62 | 100 |
| Mobile  HART | 97 | 100 | 99.8 | 99.55 | 100 |
| Proposed model | 95.97 | 100 | 99.19 | 98.34 | 100 |

We measured the clustering results of three HART models. The dataset used for clustering used 1 million bio-signals generated from the elderly in the past. The experimental results showed that the initial HART did not cluster well. MobileHART had some problems, but most clusters were formed well. Finally, the proposed algorithm showed the perfect cluster formation. Figure 4 shows the clustering results of the three models. In this figure, red and yellow indicate emergency and critical situations, respectively. Other colors indicate normal and standby conditions. Only when red and yellow are clustered well can we judge that the AI ​​model has learned well.



1. Result of dataset clustering.

# Conclusion

We proposed a service and an AI model that allows the elderly to quickly take action in emergency situations. We implemented a part that judges the elderly's falls using the latest two models and transmits data to hospitals and emergency facilities. In the service implementation, we built the infrastructure using MS Azure cloud and deployed a container-based emergency situation detection service. We compared and analyzed the performance of the proposed model with the two existing models. The evaluation results showed that the proposed model improved by approximately 15% compared to the existing model. In the future, we will further evolve the emergency situation judgment model and develop a model that can be applied to various emergency environments.

##### Acknowledgment

This work was supported by the Human IT Education Center, Seoul, Republic of Korea.

##### References

1. OECD Health Policy Studies. Is Care Affordable for Older People?, 2024.
2. P. Kumari, L. Mathew, and P. Syal. Increasing trend of wearables and multimodal interface for human activity monitoring: A review, Biosensors Bioelectron., vol. 90, pp. 298307, 2017.
3. D. H. Sim and H. U. Yoon, Monitoring a Driver's Respiration Status by Using Seat Belt Equipped with Inertial Measurement Unit, *Journal of Korean Institute of Intelligent Systems*, vol. 32, no. 3, pp. 238–243, 2022.
4. Kwak, Y., Kim, J., & Lee, J. Transformer-based arrhythmia detection model robust to ECG noise. *Journal of Biomedical Signal Processing*, *35*(4), 215-230, 2022.
5. Jae-hoon Jeong, & Hyunho Park. Development of a False Alarm Classification and Prediction Model Using Transformer-Based Large Language Model. The Journal of Korean Institute of Communications and Information Sciences, 49(4), 489-502. 10.7840/kics.2024.49.4.489, 2024.
6. J. Park, D. Lim, S. Park, and H. Park. Transformer based Collision Detection Approach by Torque Estimation using Joint Information, Journal of Korea Robotics Society, vol. 19, no. 3. The Korea Robotics Society, pp. 266–273, 31-Aug-2024.
7. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
8. Yuan, L., Chen, Y., Wang, T., Yu, W., Shi, Y., Jiang, Z. H., ... & Yan, S. (2021). Tokens-to-token vit: Training vision transformers from scratch on imagenet. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 558-567).
9. Wang, W., Xie, E., Li, X., Fan, D. P., Song, K., Liang, D., ... & Shao, L. (2021). Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 568-578).
10. Ek, Sananra & Portet, François & Lalanda, Philippe. (2023). Transformer-based models to deal with heterogeneous environments in Human Activity Recognition. Personal and Ubiquitous Computing. 27. 1-14. 10.1007/s00779-023-01776-3.
11. Mehta, S., & Rastegari, M. (2021). Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178*.
12. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
13. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.
14. Dave, V. S., & Dutta, K. (2014). Neural network based models for software effort estimation: a review. *Artificial Intelligence Review*, *42*(2), 295-307.
15. M. Saleh, M. Abbas and R. B. Le Jeannès, "FallAllD: An Open Dataset of Human Falls and Activities of Daily Living for Classical and Deep Learning Applications," in IEEE Sensors Journal, vol. 21, no. 2, pp. 1849-1858, 15 Jan.15, 2021, doi: 10.1109/JSEN.2020.3018335.
16. Reyes-Ortiz, J., Anguita, D., Ghio, A., Oneto, L., & Parra, X. (2013). Human Activity Recognition Using Smartphones [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C54S4K
17. Blunck, H., Bhattacharya, S., Prentow, T., Kjrgaard, M., & Dey, A. (2015). Heterogeneity Activity Recognition [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5689X
18. Reiss, A. (2012). PAMAP2 Physical Activity Monitoring [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5NW2H.
19. Y. -H. Nho, J. G. Lim and D. -S. Kwon, "Cluster-Analysis-Based User-Adaptive Fall Detection Using Fusion of Heart Rate Sensor and Accelerometer in a Wearable Device," in IEEE Access, vol. 8, pp. 40389-40401, 2020, doi: 10.1109/ACCESS.2020.2969453.