# AUTOMATIC ELDERLY FALL AND UNSTABLE MOVEMENT DETECTION SYSTEM USING FRAMEWISE AND LSTM BASED VIDEO ANALYTICS ON AN EMBEDDED DEVICE


**Aniqua Nusrat Zereen [1] Anubinda Gurung [2] Amir Rajak [3]**
**Jednipat Moonrinta [4] Matthew Nelson Dailey [5] Mongkol Ekpanyapong [6]**
**Roongtiwa Vachalathiti [7] Sunee Bovonsunthonchai [8]**

Department of Information and Communication Technologies, School of Engineering and Technology, Asian Institute of Technology, Pathum Thani 12120 Thailand [1 to 5]

Department of Industrial Systems Engineering, School of Engineering and Technology, Asian Institute of Technology, Pathum Thani 12120 Thailand [6]

Musculoskeletal physical therapy management and Biomechanics of human movement and sports, Physical Therapy, Faculty of Physical Therapy, Mahidol University, Nakhon Pathom 73170 Thailand [7]

Neurological physical therapy management, Gait and balance biomechanics, Human movement analysis, Physical Therapy, Faculty of Physical Therapy Mahidol University, Nakhon Pathom 73170 Thailand [8]

Corresponding E-mail : st120069@ait.ac.th

Received Date August 5, 2021
Revised Date November 9, 2021
Accepted Date November 16, 2021

## Abstract

We introduce an edge processing device and cloud computation framework enabling activity profiling, unstable motion alerts, and fall alerts for elderly people living at home under their families' care. The system analyzes video frames captured by fixed cameras, tracking each visible person and classifying their actions into one of nine ordinary activities, a fall, or unstable movement. Alert notifications are sent to caregivers if a fall or unstable movement is detected. The system comprises an embedded device (NVIDIA TX2) and cloud-based storage and analysis infrastructure. The main modules include video-based human detection, tracking, and recognition; fall detection; activity classification; and detection of painful, unstable, and confused motion likely to lead to falls. The system is designed for accuracy, usability, and cost. Individual module tests and a field test with a volunteer household indicate that the prototype system is ready for the next stage of commercial exploitation, with an accuracy of 91.6% for ordinary actions and falls as well as a recall of 97.02% for unstable motion.

**Keywords:** Automatic elder care, Activity profiling, Fall warning, Unstable motion

## 1. Introduction

Most elderly people prefer living at home, but living at home without full-time assistance can be risky due to the possibility of falls; more than one-third of persons 65 years and over fall each year, and in half of these cases, the falls are recurrent (Al-Aama, 2011). Most current research and development in video processing for elder care focuses on detecting falls and recording activities of older adults over time. On the other hand, real-time unusual behavior alerts may help predict and prevent accidents rather than speeding up a reaction after the fact. As of yet, video technology has not been developed to its full potential in this area; to our knowledge, there are no commercial video analytic products able to monitor elders' behavior and stability in real-time and inform family members or caregivers when problems arise. In this paper, we introduce a prototype Internet of Things (IoT) video processing device that not only implements basic human detection, tracking, and recognition modules for fall detection and activity summarization, but one that is also capable of monitoring elderly people's behavior and raising alerts when unusual or unstable behavior ensues. The goals of the research are improvements in video technology for elder care, an increase in the longevity of in-home care for the elderly, and increased entrepreneurial activity in the exploitation of video technology for elder care. Preliminary field testing of the prototype device proves the concept.

Research shows that delaying assistance for elders at home can lead to complications that trigger further loss of mobility and independence. Toward automated fall detection and prevention, advances in wearable sensors include commoditized wrist-wearable sensors and new innovative wearable devices such as fall prevention shoes with embedded cameras and laser-based obstacle detection (Lin et al., 2017). Wearable devices, however, are very vulnerable to noise in the environment and can be inconvenient and intrusive to the elderly (Mubashir et al., 2013). There has also been a great deal of research on the use of fixed optical sensors (Anderson et al., 2006), including analysis of movement and silhouette information over time (Zhou et al., 2009) and extraction of high-level human activity information over time (Chung & Liu, 2008; Zhou et al., 2008). Specifically, in the study by Zhou et al. (2009), accurate human silhouette extraction was carried out using various post-processing techniques from foreground segmentation, morphological filtering and continuous background update, since the changes in lighting conditions pose a significant challenge. While their approach used silhouettes that take the form of blobs, we make use of skeleton key points that provide high-level information about body pose which simplifies the process of learning action recognition since skeleton-based features are more discriminative. Depth cameras have also been used to more precisely measure spatial and temporal gait parameters (Stone & Skubic, 2011). Depth cameras are relatively large in size, more expensive than RGB cameras, and have the limited range. Wang et al. (2012) propose a new feature, the local occupancy pattern, to represent "depth appearance," which captures relationships between human body parts and environmental objects in a scene. They use the actionlet, specific conjunction of features for a subset of the joints, along with data mining, to discover the most discriminative actionlets for human action classification. They use depth cameras to capture depth map sequences from which 3D joint positions and then the local occupancy pattern feature are extracted. Compared to them, our system is more simple using multiple RGB cameras for different viewpoints and then training on multi-view 2D skeleton data to provide view invariance. In contrast, Li et al. (2017) propose a skeleton-based CNN with a transformer module that can rearrange and select important joints automatically. They adopt a maxout scheme for multiple people. Our system can handle predictions for multiple people without any such scheme. Moreover, they can only predict actions in a batch processing way. Moving towards recurrent neural networks, Zhang et al. (2017) make use of long short term memory (LSTM) modules trained on geometric relational features based on distances between joints and selected lines for action recognition using skeletons. The large number of hand-crafted features and neural network parameters used in this work apparently led to over-fitting.

In the following section, we describe a complete prototype elderly activity profiling and mobility analysis system built through the integration of several existing modules and new modules that are sufficiently accurate for commercial exploitation that makes effective use of the hardware available.

## 2. Proposed Methods

### 2.1 Overview

The prototype elder care system analyzes video frames captured by fixed cameras. It tracks each visible person and classifies their actions into one of nine ordinary activities, a fall, or unstable movement. An unstable movement is considered to be the state before/after fall due to collapsing movement, heart attack, slip, hit obstacle, or dizziness. If a fall or unstable movement is detected, the system sends an alert notification to the caregiver. The system comprises an embedded device (NVIDIA TX2) and cloud-based storage and analysis infrastructure. The main modules of the system are video-based human detection, tracking, and recognition, fall detection, and detection of painful, unstable, and confused motion likely to lead to falls.

### 2.2 Hardware and Network Design

The system includes IP cameras, a network switch, and an NVIDIA TX2 embedded system. The system is installed with concern for privacy (bedrooms and bathrooms are not monitored, but entries into and exits from these rooms are monitored). Network and computer hardware are placed with the house's Internet router. All video processing is performed on premises, with the residents' permission. Information extracted from the video is forwarded to a cloud-hosted service that implements authentication and access control rules.

### 2.3 System Architecture

The system (Figure 1) comprises several modules forming an edge-cloud computing system. Edge modules include people detection and tracking. Cloud modules include face detection, face alignment, face recognition, activity classification using a new cascaded framewise activity classification module, and unstable movement analysis using a new LSTM-RNN-based unstable movement classification module. If a fall or unstable movement is detected at any time, the system generates an alert via a RESTful web service and sends a Short Message Service (SMS) to the family member or doctor. The system also sends an alert if it detects the absence of people for a specific period of time. Other activities are simply stored in the database and summarized in the individual's activity profile.

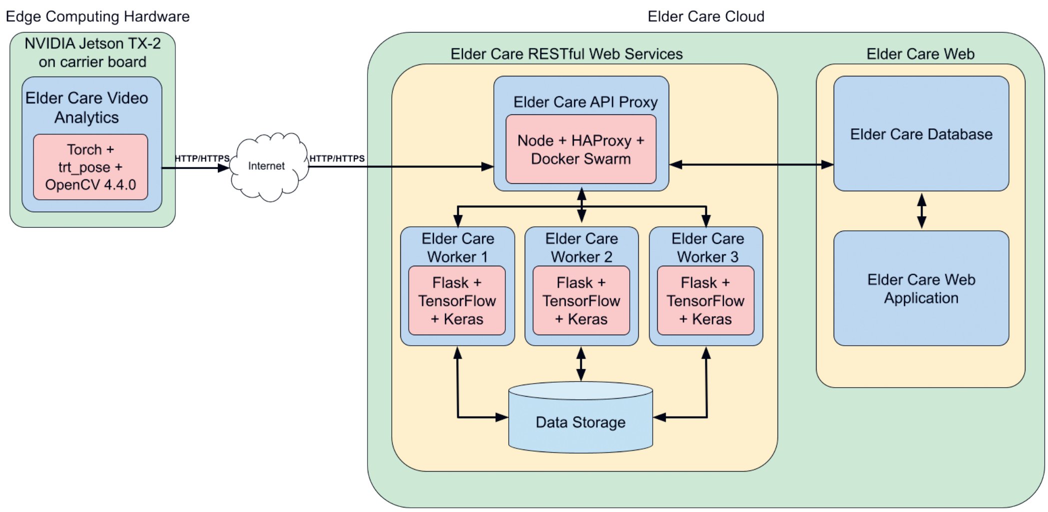



**Figure 1** System Architecture

### 2.4 Training and Evaluation Data

Researchers have proposed several categories of relevant daily physical, recreational, and communication activities for elderly people detectable through hand gestures and body postures (Chung et al., 2017; Pumpinyo & Koocharoenprasit, 2020; Yen & Lin, 2018). We have selected sit, stand, walk, and bend indicating daily mobility, as well as clap, check the time, talk on the phone, point, and wave indicating daily recreational and communication activities. We, therefore, collected examples of individuals performing these 10 activities from different sources, and we added examples of simulated falls and unstable movement. The sources include:

1. The MoVi online dataset (90 people) (Ghorbani et al., 2020).
2. Artificial Intelligence Center Lab (AIC lab; 12 people).
3. Volunteer house (AIC Volunteer Set 1; three people).
4. Volunteer house (AIC Volunteer Set 2; two weeks later than Set 1; three people).

Datasets 2, 3, and 4 contain 10 activities. Dataset 1 contains nine activities (no falls).

For the new datasets (2, 3, and 4), at each location, we recorded video using eight cameras covering a 360° view and cropped them into short clips, each expressing one activity. We processed each frame of each clip to detect people and estimate their pose, then stored each resulting sequence of skeleton points with a corresponding activity label in a CSV file for training and testing.

### 2.5 People Detection, Tracking, and Recognition

The human detection system is based on OpenPose (Cao et al., 2019) or TRTPose (Xiao et al., 2018). We expand the bounding box of the detected skeleton points. Tracking is performed using DeepSORT (Wojke et al., 2017). Recognition modules include face detection and face alignment using MTCNN (Zhang et al., 2016) and face recognition using Facenet (Karim et al., 2018), which are used to detect people in each camera view and keep track of each person, accumulating regions of interest (image patches corresponding to the detected person) in consecutive frames from a video stream. The sequence of consecutive image patches is passed to the activity classification and unstable movement modules.

All three modules are integrated with a cloud-based storage and analysis infrastructure that sends a notification to caregivers when a fall, unstable movement, or an absence of people for a specified period of time by the family member is observed. The notification service is implemented via the LINE Notify API to enable ease of use and is accessible via mobile phone or computer. Users must provide a LINE Notify token key or their LINE user name during the configuration of the Web application to enable this feature. Notification messages are sent to users with an indication of the incident type (fall, unstable movement, or absence) and a URL for accessing the image corresponding to the event. URLs are hashed to prevent privacy violations.

### 2.6 Framewise Cascaded CNN for Activity Classification

The workflow of our activity classification module (Figure 2) contains a cascaded hierarchical classifier to classify 10 activities including fall. Since checking the time, talking on the phone, clapping, pointing, and waving might occur contemporaneously with sitting, standing, or walking, but not (by assumption) with bending or falling, we first differentiate the major groups of activities among the 10 activities then, if warranted, classify into the minor activities (check time, clap, phone call, point, and wave). The Activity-1 classifier performs the gross classification based on 36 skeletal keypoints, and if the gross activity is sitting, standing, or walking, we activate one of the three Activity-2 classifiers, each of which classifies the input into the fine-grained categories check time, clap, phone call, point, wave, or none of the above (only sitting, standing, or walking, without any fine-grained activity). The fine-grained Activity-2 classifiers are trained only on the upper half 24 skeletal keypoints of the human body. Whenever any of the upper half 24 skeletal keypoints are not visible, we simply output the gross category (sit, stand, or walk) without a fine-grained category. As mentioned earlier, the fall category activates a notification. The unstable movement detection module is separate from the framewise activity classifier and runs in parallel.

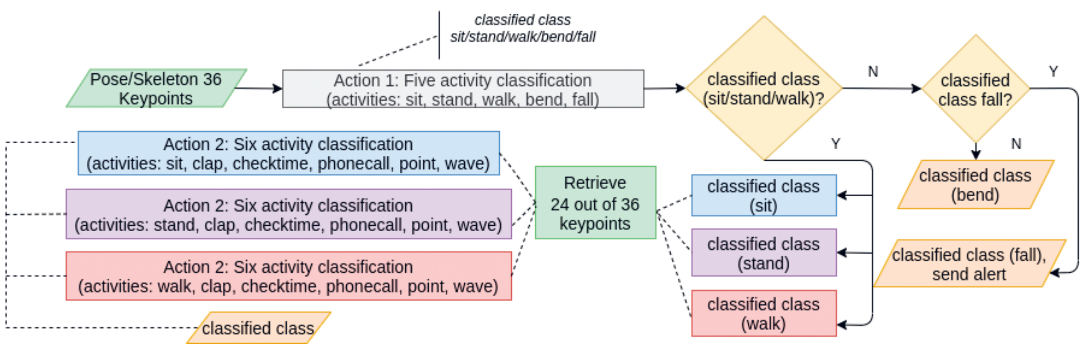



**Figure 2** Workflow of the Cascade Framewise Classifier

The Activity-1 model consists of five dense layers, each followed by batch normalization and ReLU. The output layer is a five-unit linear layer with softmax. Training minimizes cross-entropy loss on 32-frame batches with the Adam optimizer (learning rate 0.001) for 480 iterations. Class weight is balanced across samples. The three Activity-2 models are the same as the Activity-1 model except that they have six units in the output layer rather than five. The detailed model parameters are summarized in Figure 3. We combined the three datasets then split them randomly into 70% for training, 10% for validation, and 20% for the test to evaluate the final model. A summary of the training, validation, and test set distributions for the four models is provided in Table 1. Test confusion matrices after training are shown in Figure 4. The overall accuracy of the test set (considering all four models) is 96%.

| Model: "sequential_1" | Activity_1 | | Activity_2 | |
|---|---|---|---|---|
| Layer (type) | Output Shape | Param # | Output Shape | Param # |
| dense_1 (Dense) | (None, 128) | 4736 | (None, 128) | 3200 |
| batch_normalization_1 | (None, 128) | 512 | (None, 128) | 512 |
| dense_2 (Dense) | (None, 64) | 8256 | (None, 64) | 8256 |
| batch_normalization_2 | (None, 64) | 256 | (None, 64) | 256 |
| dense_3 (Dense) | (None, 32) | 2080 | (None, 32) | 2080 |
| batch_normalization_3 | (None, 32) | 128 | (None, 32) | 128 |
| dense_4 (Dense) | (None, 16) | 528 | (None, 16) | 528 |
| batch_normalization_4 | (None, 16) | 64 | (None, 16) | 64 |
| dense_5 (Dense) | (None, 8) | 136 | (None, 8) | 136 |
| batch_normalization_5 | (None, 8) | 32 | (None, 8) | 32 |
| dense_6 (Dense) | (None, 5) | 45 | (None, 6) | 54 |
| | Total params: 16,773 | | Total params: 15,246 | |
| | Trainable params: 16,277 | | Trainable params: 14,750 | |
| | Non-trainable params: 496 | | Non-trainable params: 496 | |

**Figure 3** Activity-1 and Activity-2 Model Parameters.

**Table 1** Dataset Distribution for Framewise Activity Classification Cascade; Preclass: Activity-1; Postclass: Activity-2

| Model | Train | Validation | Test | Test Accuracy |
|---|---|---|---|---|
| Preclass | 92610 | 10290 | 25725 | 98.00 |
| Postclass-sit | 78477 | 8720 | 21800 | 96.00 |
| Postclass-stand | 84481 | 9387 | 23468 | 95.00 |
| Postclass-walk | 94346 | 10483 | 26208 | 95.00 |

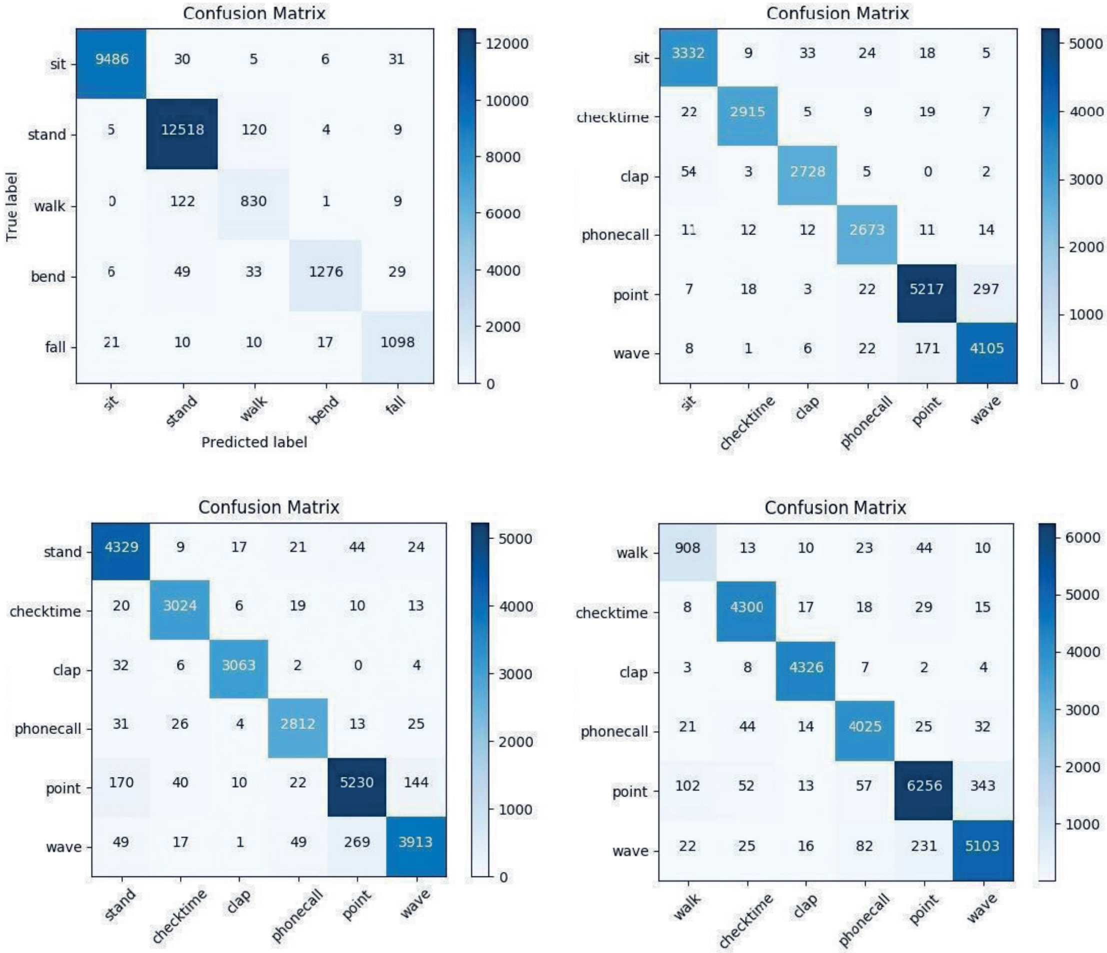



**Figure 4** Confusion Matrix for Activity Classification on Test Set
Left to right: Activity-1, Activity-2 Sit, Activity-2 Stand, and Activity-2 Walk

## 2.7 Sequence-based Stable/Unstable Movement Classification

Figure 5 shows the architecture of the RNN model used for unstable movement prediction. We use LSTM cells to capture information across a series of frames. The input to the model is a 40-frame sequence of 36 skeletal keypoints. To create the training set, we used

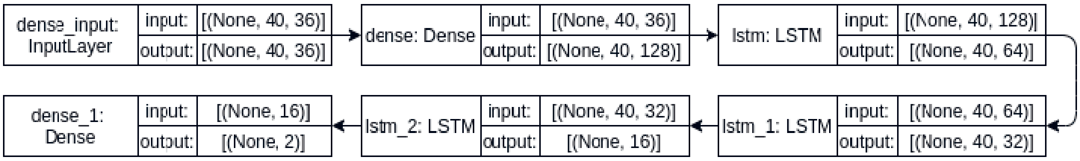



**Figure 5** Architecture of the Unstable Model

The 40 x 36 input layer is followed by a fully connected dense layer, three LSTM layers, then a final dense layer with two outputs (stable or unstable motion). To reduce overfitting and improve generalization, 40% dropout is applied to the output of the 2nd LSTM layer, 50% dropout is applied to the output of the final LSTM layer, and L2 regularization is applied throughout the model. A batch size of 2000, learning rate of 0.0005, and cross-entropy loss are used for training. We trained the model for 94 epochs with Adam optimizer.

The training, validation, and test data for the unstable motion classifier excluded Dataset 1 and the fall category examples in Datasets 2 and 3 but were otherwise the same as for the framewise activity classifier, except where there were less than 40 frames in a given sequence. Overall test set accuracy was 94.72%.

**Table 2** Stable/Unstable Motion Classifier Dataset Distribution and Accuracy

| Set | Stable | Unstable | Accuracy |
|---|---|---|---|
| Train | 182936 | 48024 | 96.06 |
| Validation | 28054 | 7095 | 94.32 |
| Test | 51682 | 13989 | 94.72 |

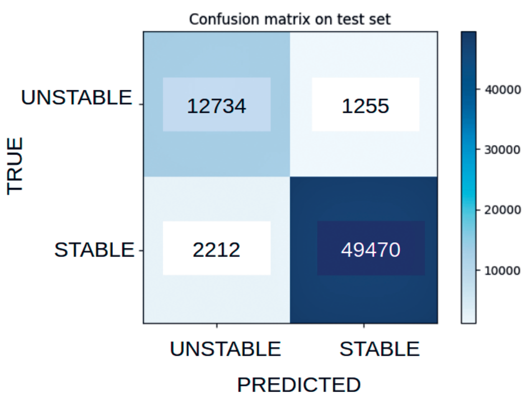



**Figure 6** Stable/Unstable Motion Classifier Confusion Matrix

## 3. Results

The final system was tested on the AIC-Volunteer-Set-2 dataset. We provide detailed results for the framewise activity classifier and the unstable motion detector.

### 3.1 Framewise Activity Classification

The final test data from the AIC-Volunteer-Set-2 dataset comprised 132 video files exhibiting the 10 activities. Table 3 shows the dataset distribution for each activity and Table 4 shows the final framewise test result for each of the models in the cascade.

**Table 3** AIC-Volunteer-Set-2 Dataset Distribution

| Sit | Stand | Walk | Bend | Fall | Clap | Check time | Point | Phone call | Wave | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| 523 | 110 | 746 | 757 | 56 | 521 | 470 | 388 | 412 | 307 | 4290 |

**Table 4** Framewise Activity Classification Results of AIC-Volunteer-Set-2

| Model | Test dataset | Accuracy |
|---|---|---|
| Preclass | 4290 | 82.00 |
| Postclass-sit | 1716 | 70.00 |
| Postclass-stand | 1015 | 84.00 |
| Postclass-walk | 2844 | 73.00 |

Framewise classification tends to be noisy, with occasional misclassification within a sequence. Figure 7 shows an example of a walking sequence in which an intermediate pose appears similar to standing still. To increase the accuracy of the framewise model, we thus aggregate each sequence of 40 frames (1.6 s) and output the model's majority vote over the sequence. This improves accuracy substantially, as shown in Table 5. Out of 19 fall videos, 18 were correctly classified, and one was misclassified as bend, resulting in 95% accuracy for fall, every activity recognized with at least 83% accuracy, and 91.6% overall accuracy. Figure 8 shows results of all nine different activities and falls in different positions which are correctly classified by the cascade framewise activity classifier.


**Figure 7** Misclassification by Framewise Classifier

**Table 5** Aggregated Framewise Activity Classification Final Test Results

| | | Sit | Stand | Walk | Bend | Fall | Clap | Check time | Point | Phone call | Wave | Total | Accuracy |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Actual | Sit | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 10 | 80.00 |
| | Stand | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 100.00 |
| | Walk | 0 | 1 | 12 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 14 | 86.00 |
| | Bend | 0 | 0 | 0 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 100.00 |
| | Fall | 0 | 0 | 0 | 1 | 18 | 0 | 0 | 0 | 0 | 0 | 19 | 95.00 |
| | Clap | 1 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 16 | 94.00 |
| | Check time | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 1 | 0 | 0 | 15 | 93.33 |
| | Point | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 1 | 15 | 93.33 |
| | Phone call | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 2 | 12 | 83.33 |
| | Wave | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 10 | 11 | 91.00 |
| Prediction | | | | | | | | | | | | 132 | 91.60 |


(a) (b) (c) (d) (e)

(f) (g) (h) (i)

(j) (k) (l) (m) (n)

(o) (p) (q)

**Figure 8** Examples of correct aggregated framewise classification results in the final test. (a – i) Nine different activities. (j – q) Falls in different positions.

### 3.2 Sequence-based Stable/Unstable Movement Classification

On the AIC-Volunteer-Set-2 dataset, the model successfully identifies unstable movements such as collapse, a simulated heart attack, dizziness, unstable movement after a fall, slipping, or hitting an obstacle, as well as stable motion. The stable/unstable motion classifier achieves an accuracy of 79.26% with a recall for unstable motion of 97.02%. Figure 9 shows the confusion matrix of test results.

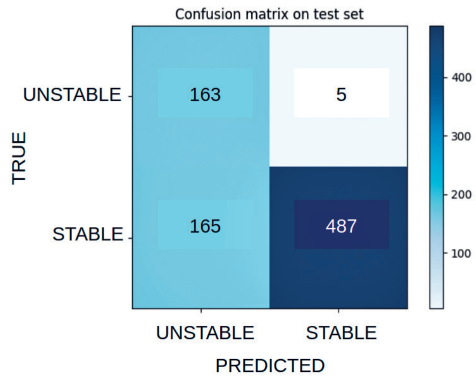
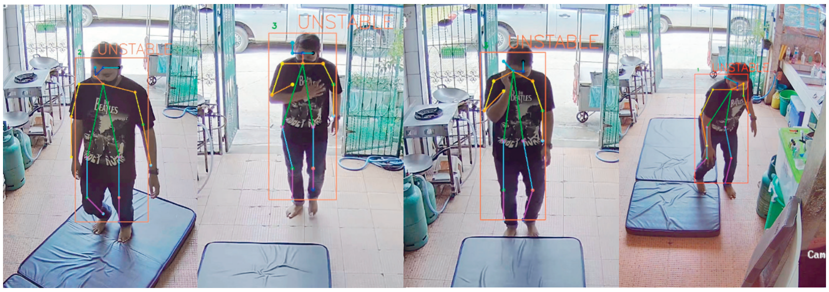



**Figure 9** Stable/Unstable Motion Classifier Final Test Results




(a) (b) (c) (d)


(e) (f) (g) (h) (i)

**Figure 10** Sample correct stable/unstable motion classification results.
(a – d) Unstable motion classification. (e – i) Stable motion classification.

## 3.3 Field Test of Integrated System

The edge-cloud integrated system includes the people identification, cascaded framewise activity classification, and unstable movement detection modules along with activity profiling. The web service displays activity summaries for individuals as well as camera views and periods of absence. Figure 11 shows sample screenshots from the web service and notifications.

**Figure 11** Sample Web Service Results and Sample Notification Alerts Corresponding Images

### 3.4 Usability Study

A survey of the overall system's usability from a health care perspective by coauthors SB and RV indicates that as the system has been designed to be easy to use and secure, as a monitoring system, it will be useful for future healthcare professionals to develop adaptations to enhance patients' mobility and functioning in conjunction with a remote physical therapy program.

## 4. Discussion

Experiments conducted thus far with the prototype system indicate its potential usefulness and ease of use for caretakers and health care professionals responsible for elderly people residing at home. Accuracy in the limited range of conditions in the field test achieves the KPIs set out for the system and is sufficient for the next phase of commercialization. The errors that do occur are mainly due to the occlusion of one family member by another or by furniture. The main limitation is the relatively high cost of high-quality IP cameras and the NVIDIA Jetson platform systems compared to the typical Thai household budget. Another limitation is sensitivity to the conditions in which the models were trained — accuracy will drop when the same models are moved to homes with different lighting and environmental conditions. Related to this issue is the veracity of the training data — there will always be a difference between simulated behavior used to train the initial model and the real-world behavior of the residents of each home the system is installed in. To address these issues, as with most deep learning models, there will be a period of fine tuning to every new site until a sufficiently wide range of conditions is encountered and incorporated into the models.

## 5. Conclusion

In this paper, we have introduced a system to enhance the care of the elderly at home. Tests of the individual modules and an overall field integration test with volunteers indicate sufficient accuracy and usability to move forward toward a commercial exploitation phase.

## 6. Suggestions

### 6.1 Suggestions for future study

The next phase of research and development will need to focus on finding the right balance between the costs and benefits of the system to elderly people and their caregivers. Since the research described here began, NVIDIA released a new embedded system board, the XAVIER NX, that is cheaper and more powerful than the Jetson TX2. We will explore the use of this system to reduce the cost and increase the effectiveness of the elder care system.

### 6.2 Suggestions for communication industry policy

One of the homes to deploy the elder care systems as described here at scale with low cost to the end user is the low bandwidth between the household and cloud providers data centers. We hope to see that bandwidth bottlenecks will be eliminated through the introduction of low-cost 5G mobile access and/or fiber to the home with high bandwidth to the data center. This would enable us to lower edge device costs and do more of the processing in the data center, where large economy of scale can be exploited.

# References

Al-Aama, T. (2011). Falls in the elderly: spectrum and prevention. *Canadian Family Physician, 57*(7), 771-776.

Anderson, D., Keller, J. M., Skubic, M., Chen, X., & He, Z. (2006). Recognizing falls from silhouettes. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 6388-6391). IEEE.

Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(1), 172-186.

Chung, J., Ozkaynak, M., & Demiris, G. (2017). Examining daily activity routines of older adults using workflow. *Journal of Biomedical Informatics, 71*, 82-90.

Chung, P. C., & Liu, C. D. (2008). A daily behavior enabled hidden Markov model for human behavior understanding. *Pattern Recognition, 41*(5), 1572-1580.

Ghorbani, S., Mahdaviani, K., Thaler, A., Kording, K., Cook, D. J., Blohm, G., & Troje, N. F. (2020). Movi: *A large multipurpose motion and video dataset*. arXiv preprint. https://arxiv.org/abs/2003.01888

Karim, N. T., Jain, S., Moonrinta, J., Dailey, M. N., & Ekpanyapong, M. (2018). Customer and target individual face analysis for retail analytics. In *2018 International Workshop on Advanced Image Technology (IWAIT)* (pp. 1-4). IEEE.

Li, C., Zhong, Q., Xie, D., & Pu, S. (2017). Skeleton-based action recognition with convolutional neural networks. In 2017 *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 597-600). IEEE.

Lin, T. H., Yang, C. Y., & Shih, W. P. (2017). Fall prevention shoes using camera-based line-laser obstacle detection system. *Journal of Healthcare Engineering*. Article 8264071.

Mubashir, M., Shao, L., & Seed, L. (2013). A survey on fall detection: Principles and approaches. *Neurocomputing, 100*, 144-152.

Pumpinyo, S., & Koocharoenprasit, S. (2020). A Survey of Leisure Activities that the Elderly Desire. *Journal of Advanced Research in Social Sciences, 3*(3), 14-19.

Stone, E. E., & Skubic, M. (2011). Evaluation of an inexpensive depth camera for passive in-home fall risk assessment. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health) and Workshops* (pp. 71-77). IEEE.

Wang, J., Liu, Z., Wu, Y., & Yuan, J. (2012). Mining actionlet ensemble for action recognition with depth cameras. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1290-1297). IEEE.

Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 3645-3649). IEEE.

Xiao, B., Wu, H., & Wei, Y. (2018). Simple baselines for human pose estimation and tracking. *In Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 466-481).

Yen, H. Y., & Lin, L. J. (2018). Quality of life in older adults: Benefits from the productive engagement in physical activity. *Journal of Exercise Science & Fitness, 16*(2), 49-54.

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters, 23*(10), 1499-1503.

Zhang, S., Liu, X., & Xiao, J. (2017). On geometric features for skeleton-based action recognition using multilayer lstm networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 148-157). IEEE.

Zhou, Z., Chen, X., Chung, Y. C., He, Z., Han, T. X., & Keller, J. M. (2008). Activity analysis, summarization, and visualization for indoor human activity monitoring. *IEEE Transactions on Circuits and Systems for Video Technology, 18*(11), 1489-1498.

Zhou, Z., Dai, W., Eggert, J., Giger, J. T., Keller, J., Rantz, M., & He, Z. (2009). A real-time system for in-home activity monitoring of elders. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 6115-6118). IEEE.