SOON CHUN HYANG UNIVERSITY

2024 SAIL Seminar

FL-FD: Federated learning-based fall detection with multimodal data fusion

Pian Qi, Diletta Chiaro, Francesco Piccialli*

Information fusion (2023.11)

SAIL

Senseable Artificial Intelligence Laboratory

순천향대학교 미래융합기술학과

Senseable AI Lab

석사과정 김병훈

Index
Contents

1 Introduction

1. Introduction

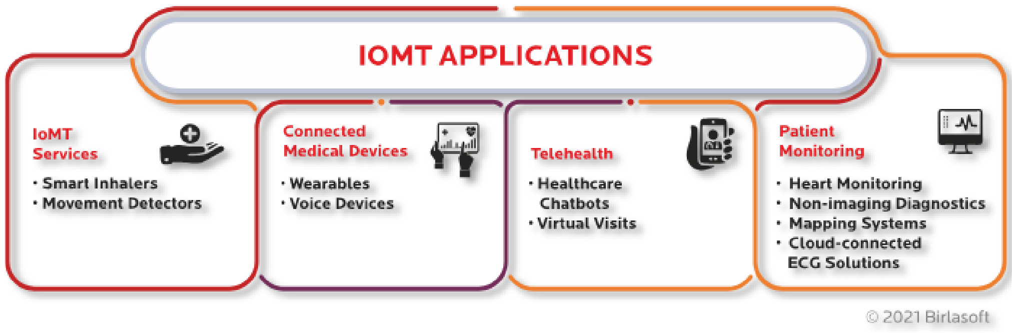
Background (Multi-modal learning)

IoMT(Internet of Medical Things)

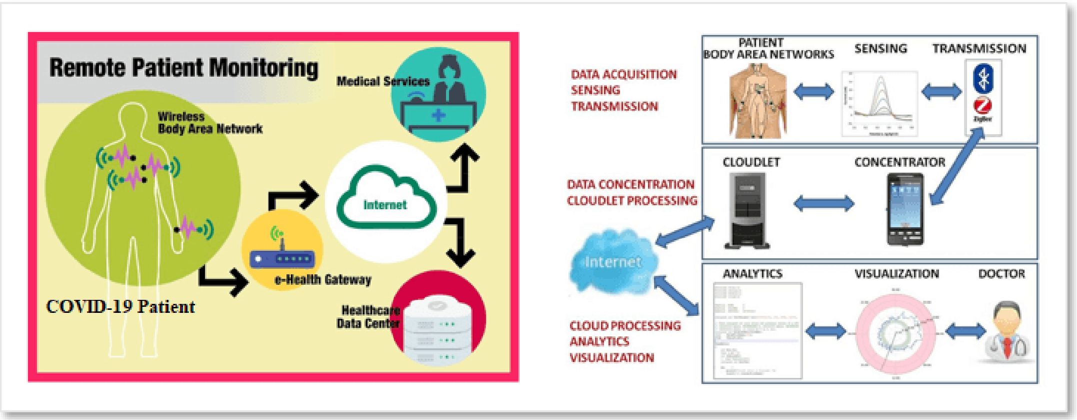
- Obtaining sensor information through wearables, cameras, and other devices
- A technology that offers many benefits to people's health and safety

Machine Learning(ML)

- Capture hidden relationships between data
- ML + IoMT : Provides support for disease prevention and diagnosis, Contributes significantly to medical advancement



<IoMT Example>



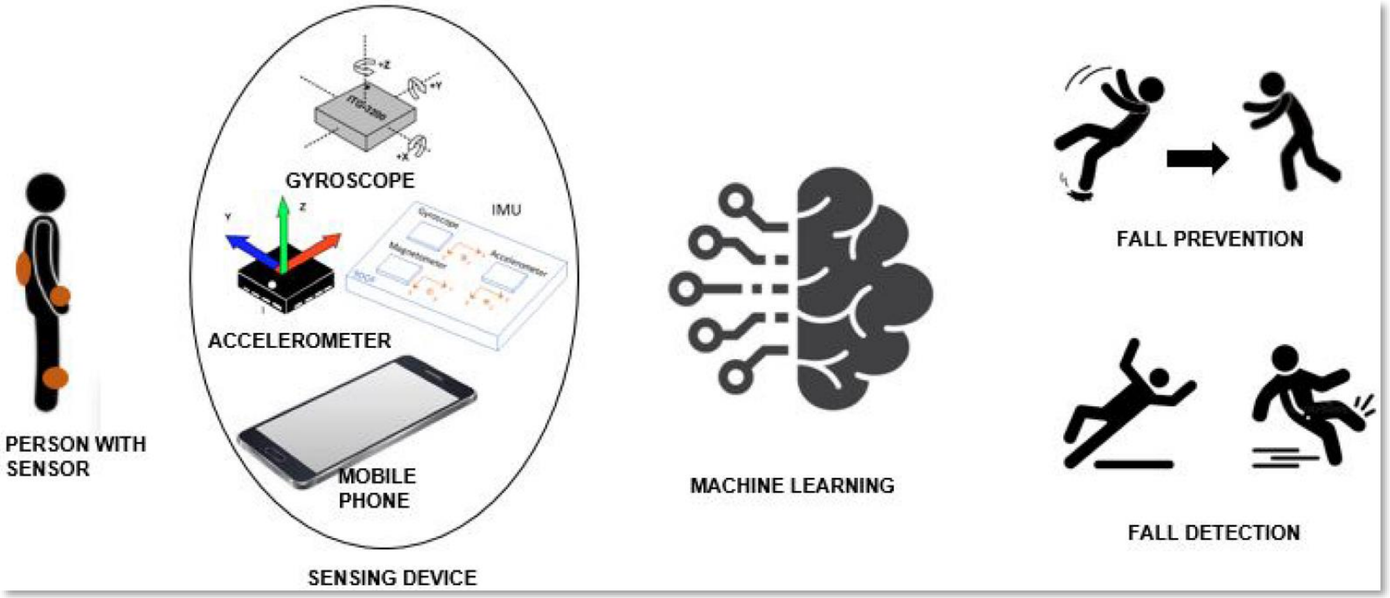
<ML Example>

1. Introduction

Background (Multi-modal learning)

Fall detection

- Falling can have serious results for vulnerable groups such as the elderly, children, and others
- It is important to detect and alert on falls
- Need to develop a fall detection system using ML + IoMT as above



<Fall detection system Example>

1. Introduction

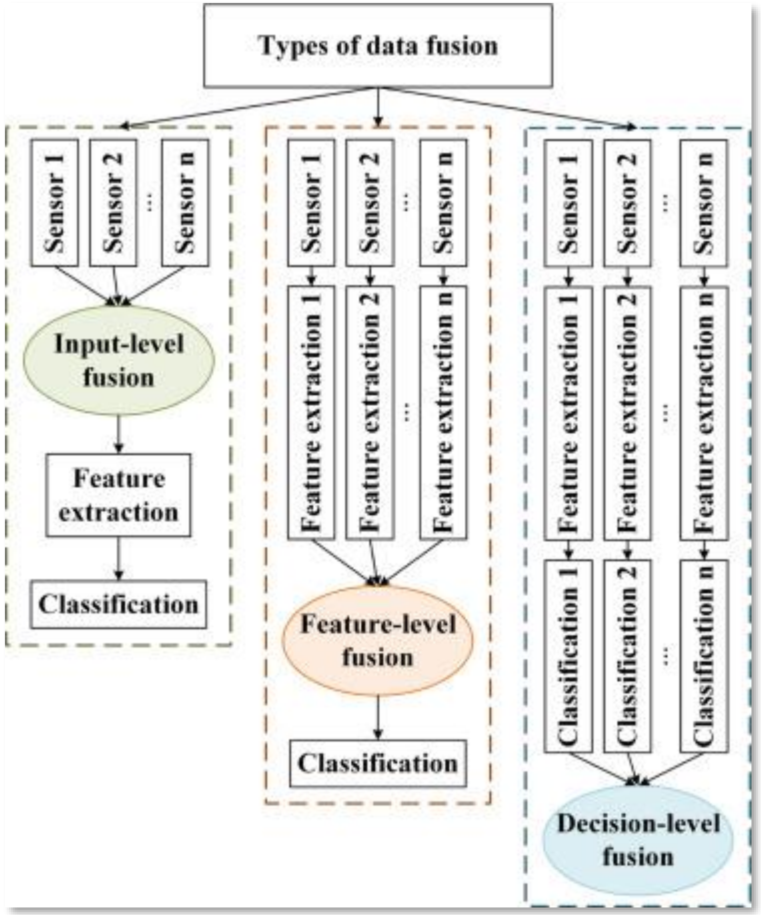
Background (Multi-modal learning)

Fall Detection System

- Rely on mostly single-modal data
- Single-modal data contains limited information
- Multi-modal data-driven systems help improve performance and can complement each other

Types of data fusion

- Input-level fusion
- Feature-level fusion
- Decision-level fusion (Ensemble)



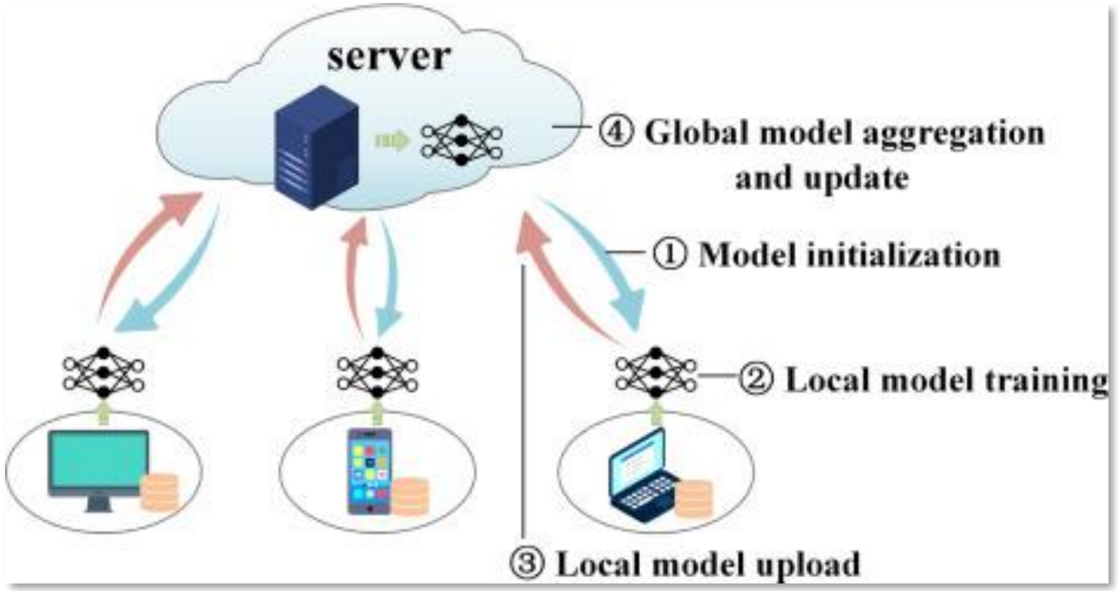
<Data fusion Example>

1. Introduction

Background (Federated learning)

Federated learning

- Centralized ML collects large amounts of user data on a central server to train models
- Privacy can be protected by using federated learning that does not share data



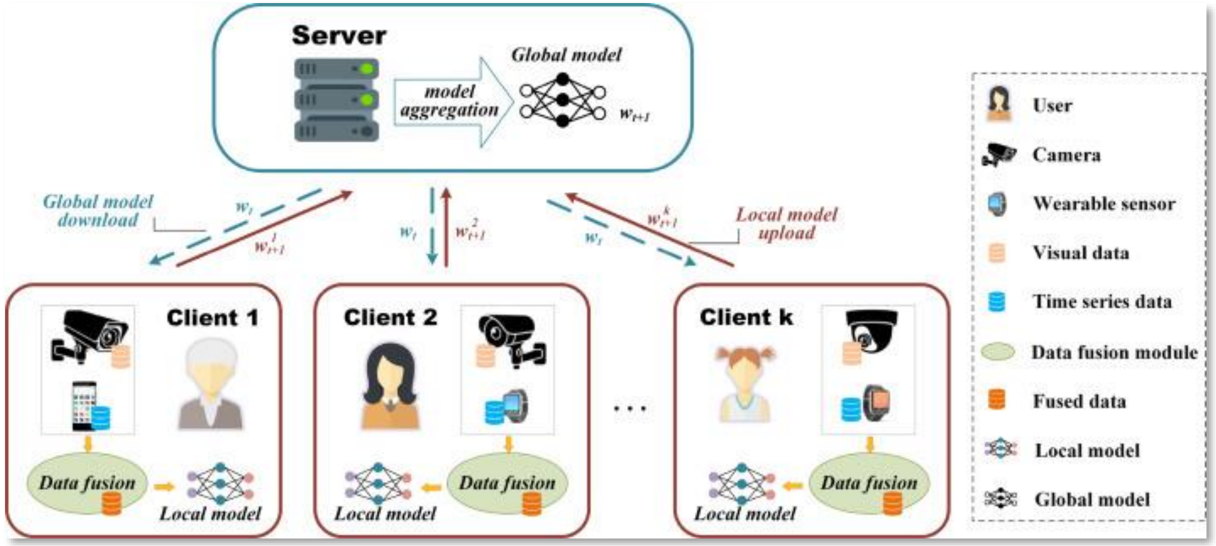
<Federated learning Example>

1. Introduction

Contribution

Contribution

1. Proposed an input-level data fusion method to combine one-dimensional time series data and two-dimensional visual data to achieve information complementarity
2. Proposed a FL framework that protects user privacy to ensure data security
3. Used benchmark dataset UP-Fall to measure efficiency, improving fall detection performance when compared to single-modal data



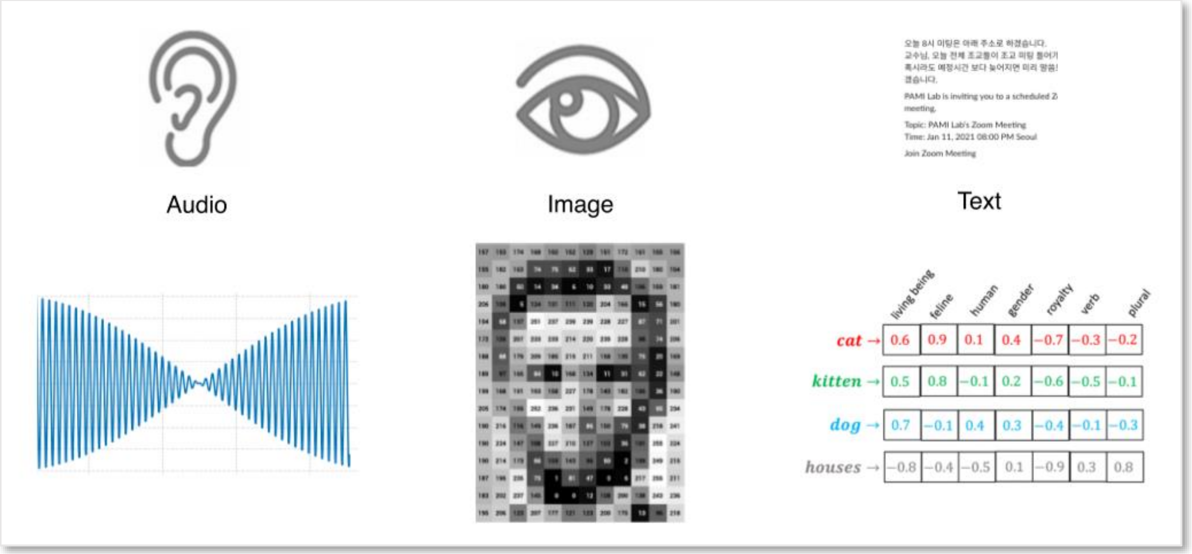
<Proposed Framework>

2 Related works

2. Related works

Data fusion in fall detection

Input-level Fusion



<Input-level Fusion Example>

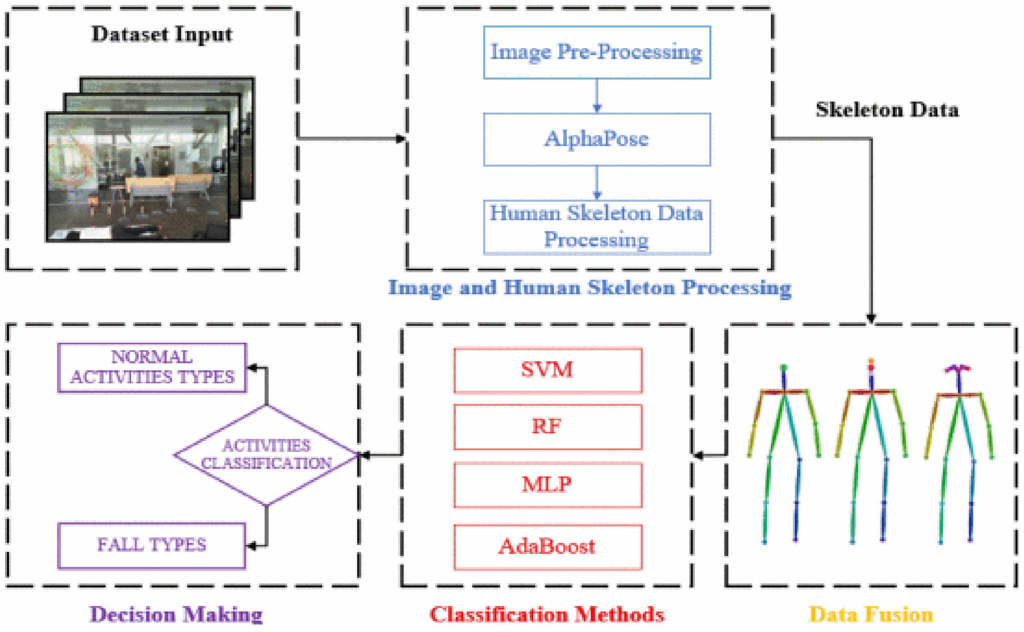
- Input-level fusion helps fully maintain information from different modalities of data and improve accuracy
- Previous work on fall detection, Input-level Fusion methods were based on single-modal data, which is limited by the heterogeneity of the data.

2. Related works

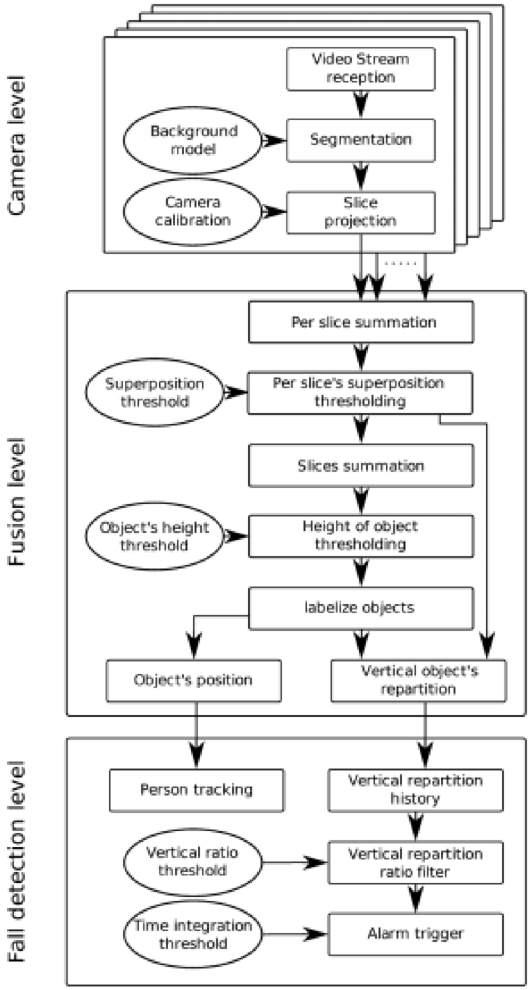
Data fusion in fall detection

Input-level Fusion with single-modal data

- Xie et al. [19] : Body skeleton + Face skeleton fusion
- Auvinet et al. [20] : Multi-camera view fusion



<Skeleton fusion framework>



<Multiple Cameras framework>

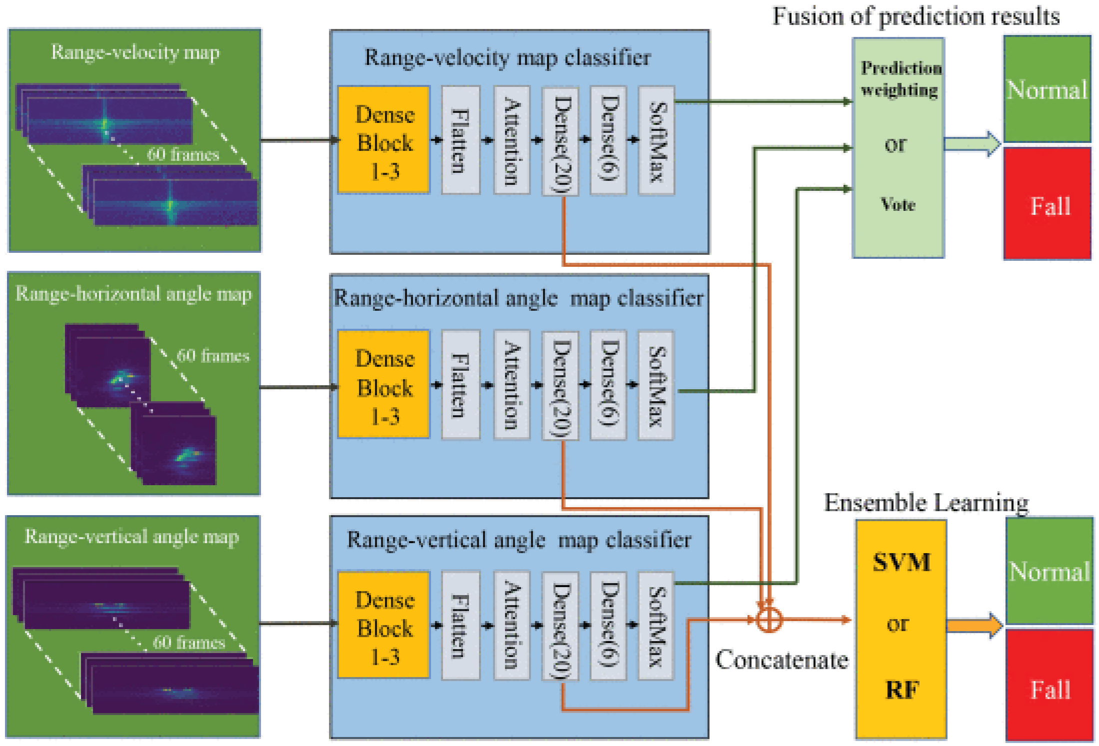
Multimodal considerations X

2. Related works

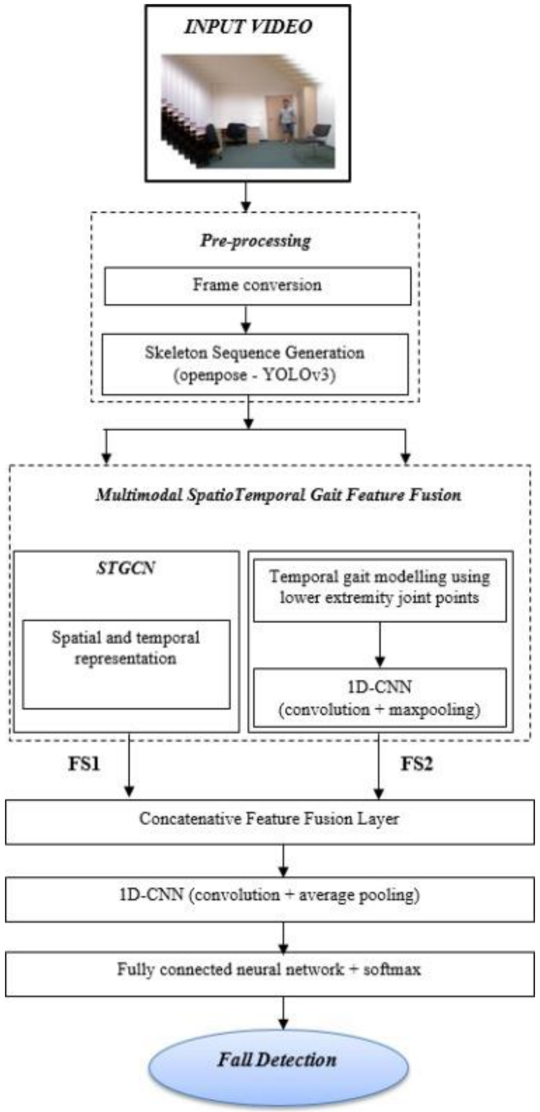
Data fusion in fall detection

Feature-level Fusion

- Yao et al. [25] : three NNs fusion
- Amsaprabhaaet et al. [26] : ST-GCN + 1D-CNN fusion



<Three NNs framework>



<ST-GCN + 1D-CNN framework>

Multimodal considerations X

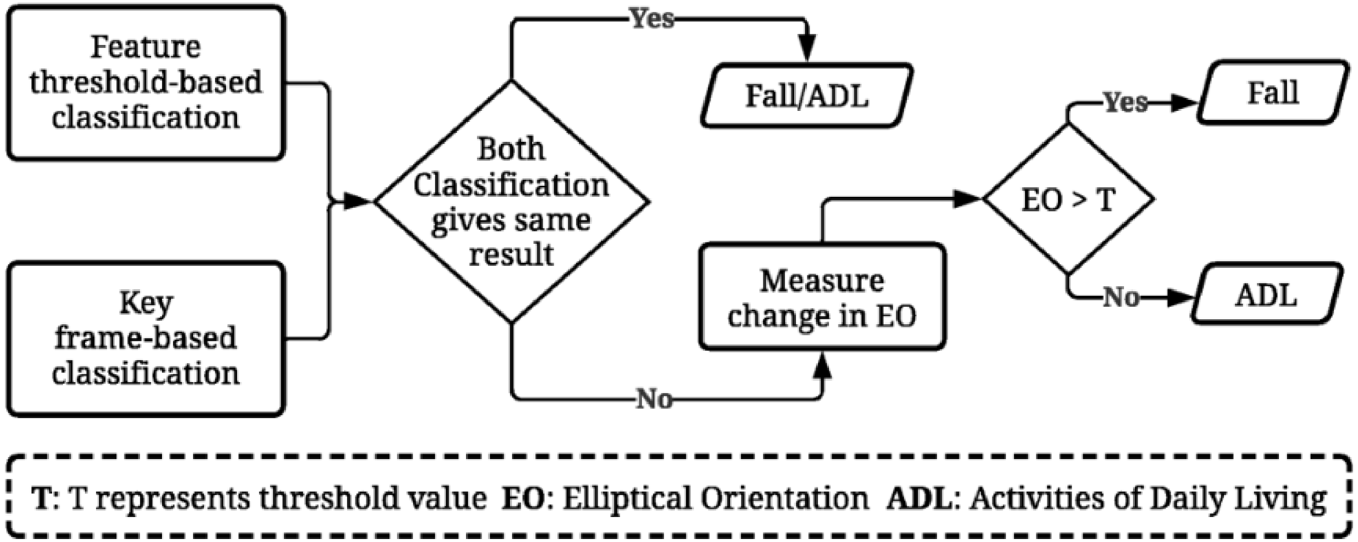
Fall-like behavior is classified as a fall

2. Related works

Data fusion in fall detection

Decision-level Fusion

- Data from different single modals is analyzed separately and then classification results are distinguished
- This approach is unable to utilize complementary information from different modal data



2. Related works

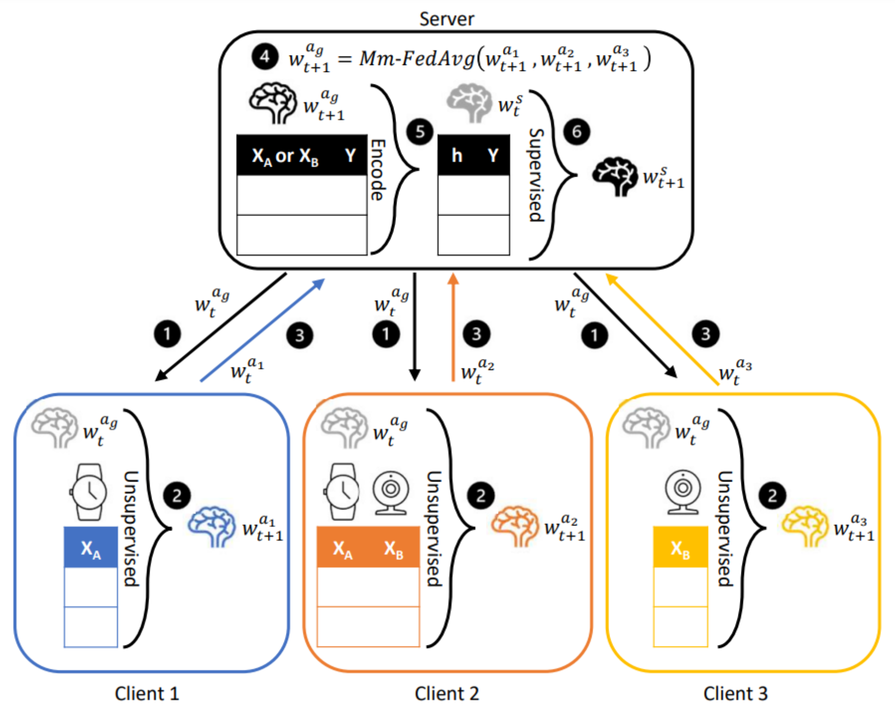
FL in fall detection

Fall detection using FL

1. Privacy-preserving federated learning based on multi-key homomorphic encryption (2022)
    - Contribution : Communication consumption efficiency, Protect privacy
    - Limitation : Not considering Non-IIDs, Single-modal data

※ Non-IID(Independent and Identically Distributed)

2. Multimodal federated learning on IoT data (2022)
    - Contribution : Multimodal data processing
    - Limitation : Communication overhead



2. Related works

Summary table

Table 1
Literature summary table.
The following table reports the main findings of the reviewed literature relating data fusion and FL in fall detection.

Papers	Multimodal	Data fusion	FL	Contributions	Limitations
[19,20]	✗	✓	✗	Input-level data fusion provides convenience subsequent fall detection tasks.	The data used for fusion have the same modality and the systems lack users' privacy protection.
[21,22,22–24]	✓	✓	✗	Feature-level data fusion helps for improving the accuracy of fall detection.	The features are manually extracted, and how to choose depends on the experience of the researcher. Lack of users privacy protection.
[25,26]	✗	✓	✗	Feature-level data fusion helps for improving the accuracy of fall detection.	The features used for fusion come from the same modality and the systems lack users' privacy protection.
[28]	✓	✓	✗	Data Fusion at decision Level.	Lack of users' privacy protection.
[31–35]	✗	✗	✓	The importance of protecting user privacy is considered, and FL is utilized to achieve the purpose.	The data used are single modality and could not provide complementary information.
[36]	✓	✗	✓	Representation of multimodal data is realized in FL, and user privacy is protected.	Data fusion was not achieved and could not provide complementary information.
Ours	✓	✓	✓	Multimodal data fusion at the input level provides information complementation, and the FL-based system effectively protects users' privacy.	

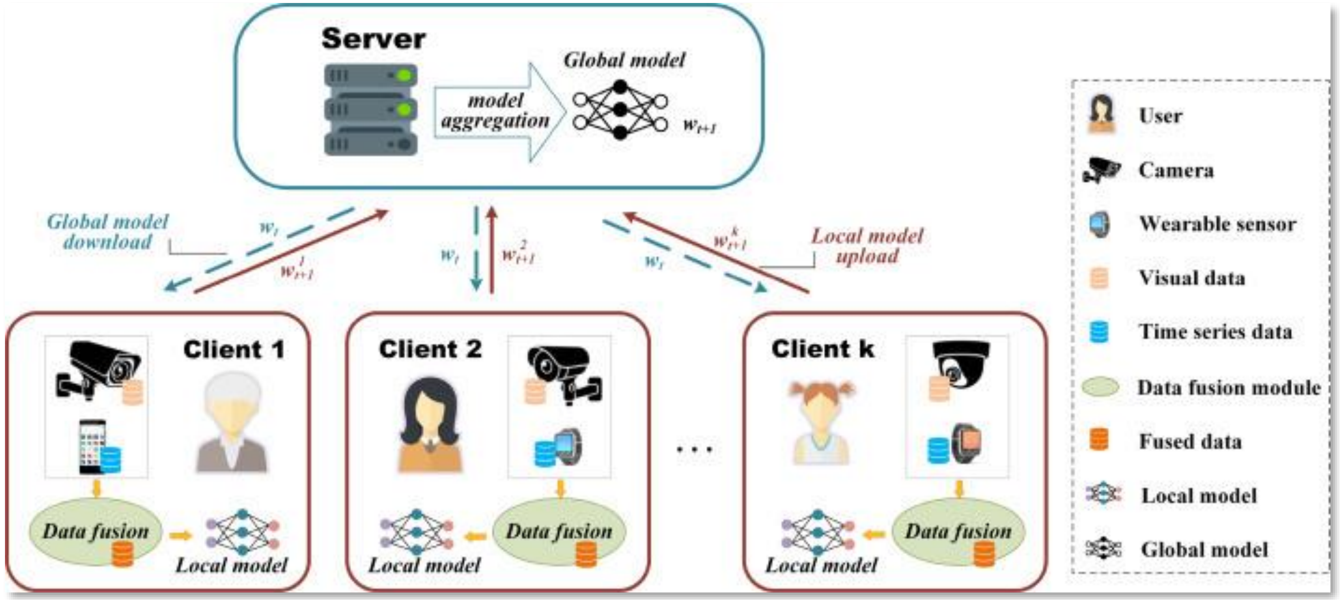
3 Proposed framework

3. Proposed framework

Overview

Proposed Framework

1. Download the global model from the server
2. Client utilize data fusion module to combine multi-modal data
3. Upload local model parameters to the server
4. Aggregate the obtained local models and update the global model



3. Proposed framework

Multimodal data fusion

Used Data Types

1. Sensor : One-dimensional time series data
2. Camera : Two-dimensional image data

Method for data fusion

1. Encoding sensor data into 2D images
2. Preprocessing RGB Images Using the Stacking Approach
3. Fusion of images into 3-channel data

3. Proposed framework

Multimodal data fusion

Encoding of time series data

- GAF(Gramian Angular field)
  - Represent temporal correlations at each point in time

1. Normalization

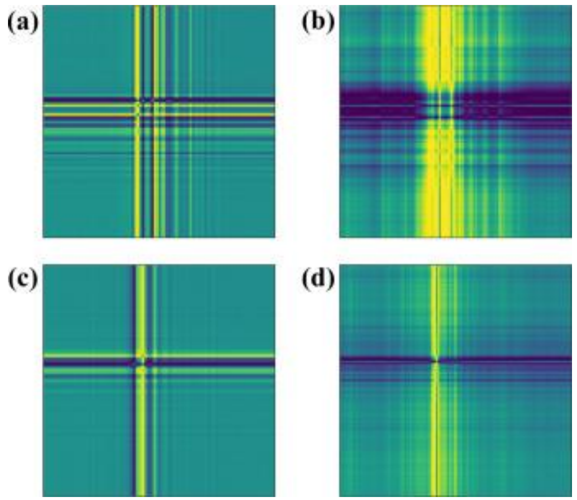
$$\widetilde{X}_T = \{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n\}$$

2. Converting spatial coordinate values to polar coordinate values

$$\begin{cases} \phi = \arccos(X_{norm,i}); \phi \in [0, \pi], \\ r = \frac{n}{L}; r \in \mathbb{R}^+, \end{cases}$$

3. Represent temporal correlations as GAF

$$GAF = \begin{pmatrix} \cos(\phi_1+\phi_1) & \cos(\phi_1+\phi_2) & \cdots & \cos(\phi_1+\phi_N) \\ \cos(\phi_2+\phi_1) & \cos(\phi_2+\phi_2) & \cdots & \cos(\phi_2+\phi_N) \\ \vdots & \vdots & \ddots & \vdots \\ \cos(\phi_N+\phi_1) & \cos(\phi_N+\phi_2) & \cdots & \cos(\phi_N+\phi_N) \end{pmatrix}$$



GAF images of different signals in UP Fall dataset, (a) Ankle accelerometer, (b) Ankle angular velocity, (c) Belt accelerometer, (d) Belt angular velocity

3. Proposed framework

Multimodal data fusion

Preprocessing RGB Images Using the Stacking Approach

- Data on falls includes both standing and lying images, so using a stacking approach to include temporal information

1. Obtain the neighboring frames

2. Calculate differences between neighboring frames

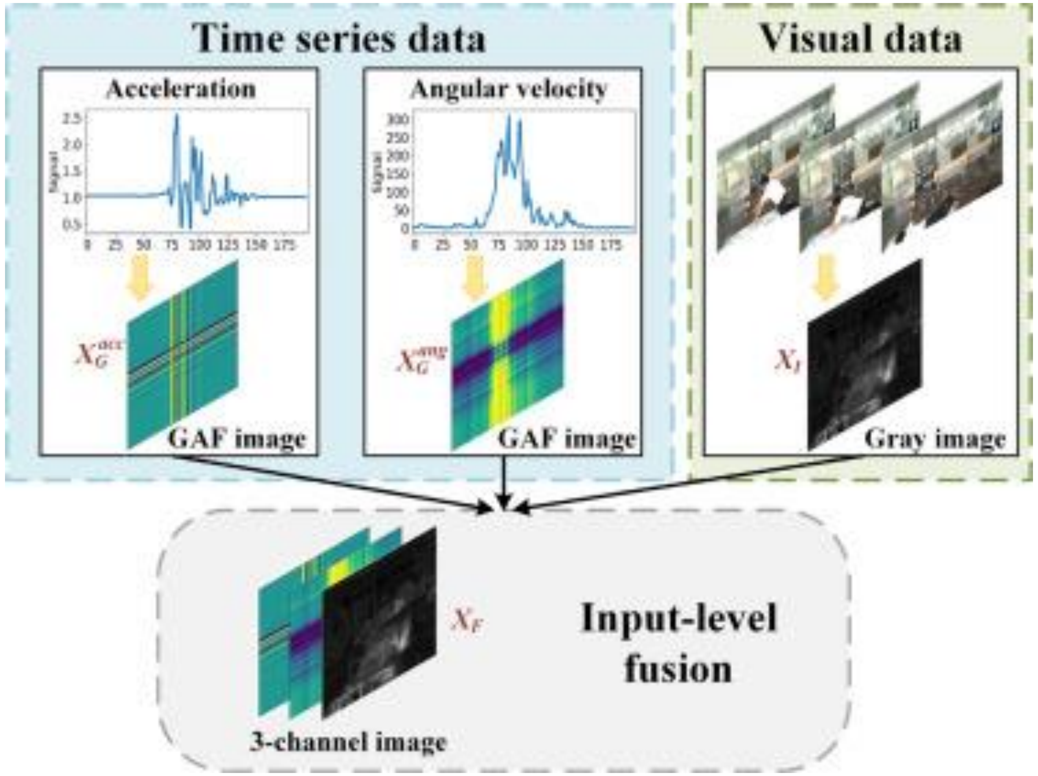
3. Convert to grayscale images

3. Proposed framework

Multimodal data fusion

Fusion of images into 3-channel data

- Acceleration GAF + Angular velocity GAF + Gray scale Image



3. Proposed framework

Federated learning deployment

FedAvg. Algorithm

- K : Number of communicating clients
- B : Local batch size
- E : Number of epochs

Algorithm 1: FedAvg algorithm.

Require: $D = (X_F, Y)$: labeled local dataset; K: number of choosed clients; B: local batch size; η: learning rate.

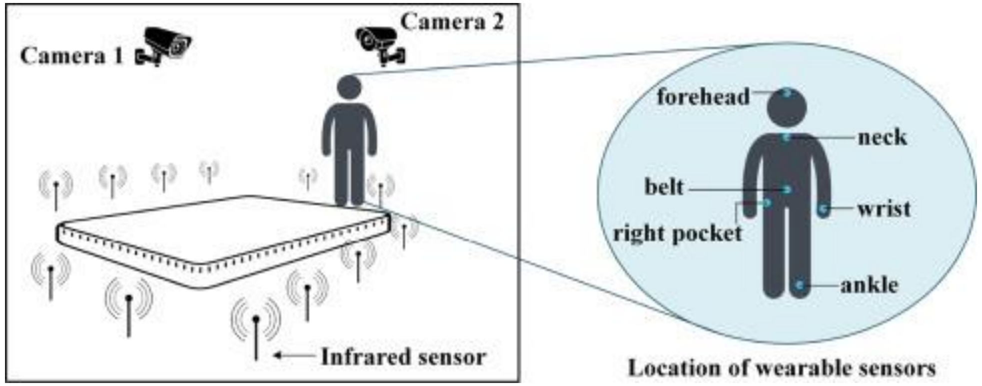
initialize w_0 at $t = 0$
for each round t **do**
$S_t \leftarrow$ randomly selected K clients
  for all client $k \in S_t$ **do**
    split D into batches of size B
    for each batch $b \in B$ **do**
      $w_{t+1}^k \leftarrow w_t - \eta \nabla \ell(w_t; b)$
    end for
  upload w_{t+1}^k to server
  end for
Server executes: $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{1}{S_t} w_{t+1}^k$
end for

4 Evaluation experiment

4. Evaluation experiment

Datasets

UP-Fall dataset



4. Evaluation experiment

Experimental design

Scenarios setup

1. Binary Classification of Falls (Falls-Non-Falls)
2. Fall activity recognition

Input data

- TS + C1 : Fusion data of time series(sensor) and camera 1
- TS + C2 : Fusion data of time series(sensor) and camera 2

Proposed

- TS : Time series only
- C1 : Camera 1 only
- C2 : Camera 2 only

4. Evaluation experiment

Experimental design

Federated Learning setup

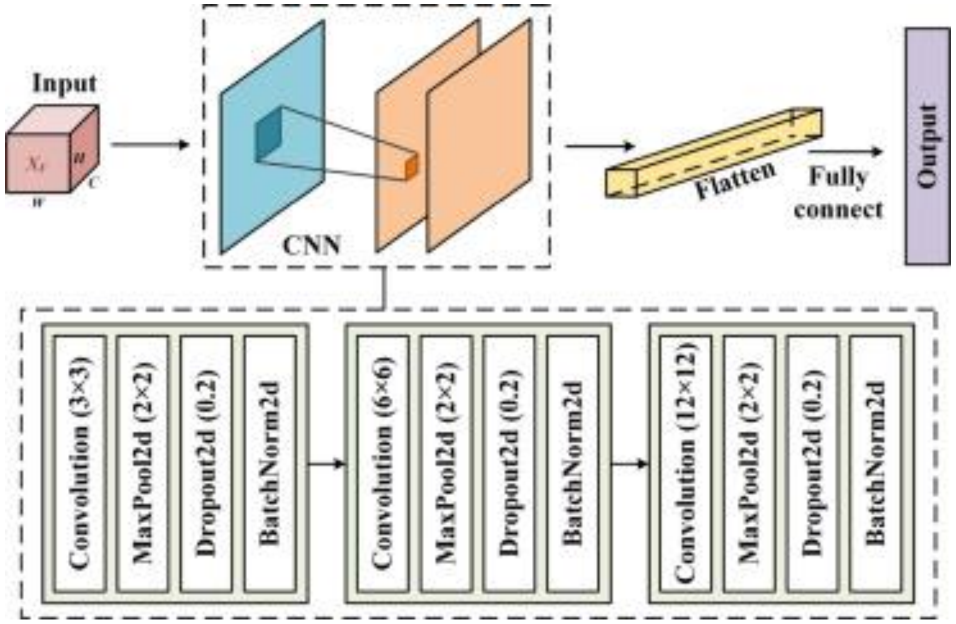
1. Round : 100(binary) / 200(action)

Model

- CNN

Evaluation Metric

- Accuracy
- Precision
- Recall
- F1-score



4. Evaluation experiment

Fall detection

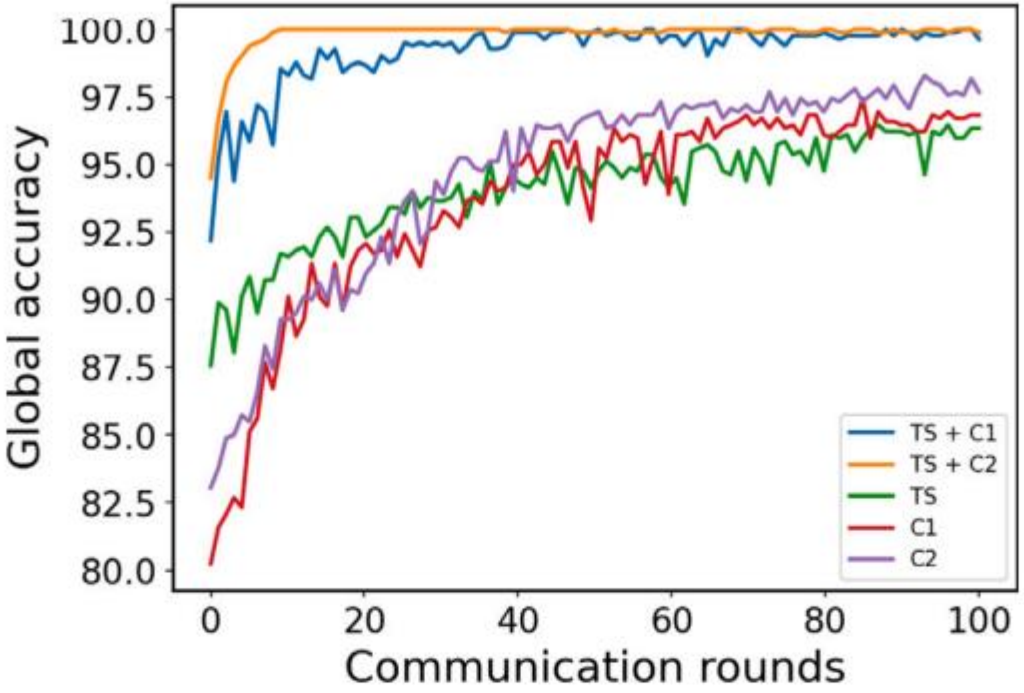


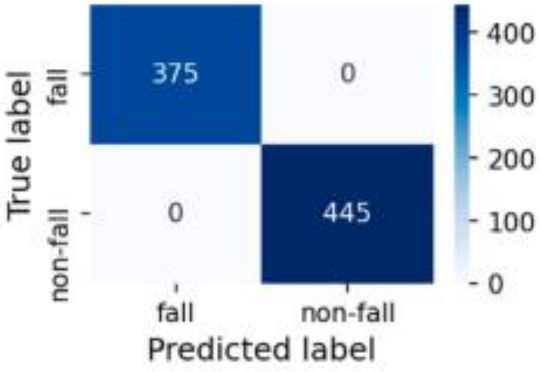
Fig. 10. The accuracy curves of global model for fall detection.

Table 3
Accuracy, Precision, Recall, and F1-score for fall detection.

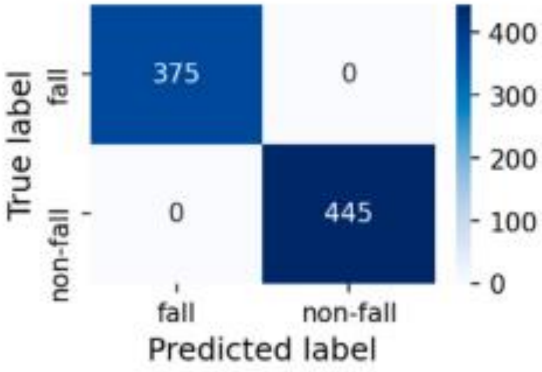
	Accuracy	Precision	Recall	F1
TS + C1	99.867 ± 0.049	99.840 ± 0.070	99.867 ± 0.071	99.854 ± 0.054
TS + C2	99.927 ± 0.041	99.841 ± 0.089	100.00 ± 0.000	99.920 ± 0.045
TS	95.880 ± 0.138	93.938 ± 0.313	97.279 ± 0.293	95.573 ± 0.147
C1	96.147 ± 0.225	94.496 ± 0.369	97.251 ± 0.297	95.847 ± 0.242
C2	97.561 ± 0.197	96.999 ± 0.385	97.706 ± 0.281	97.346 ± 0.210

4. Evaluation experiment

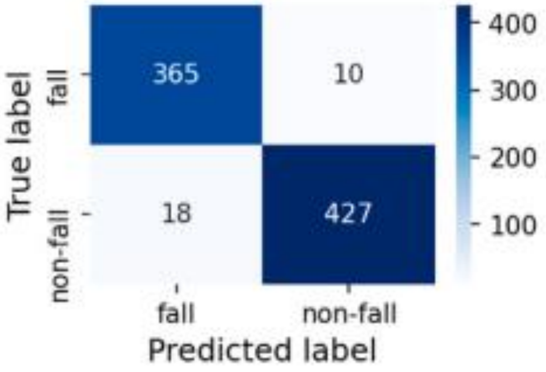
Fall detection



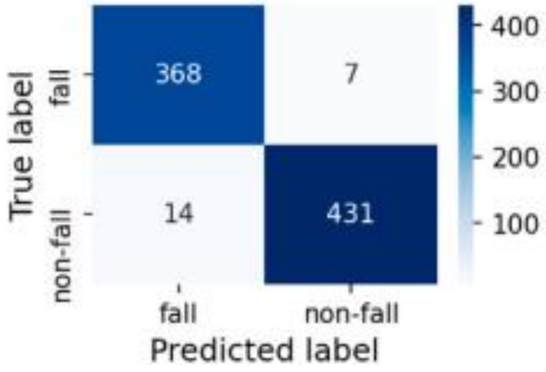
(a) TS + C1



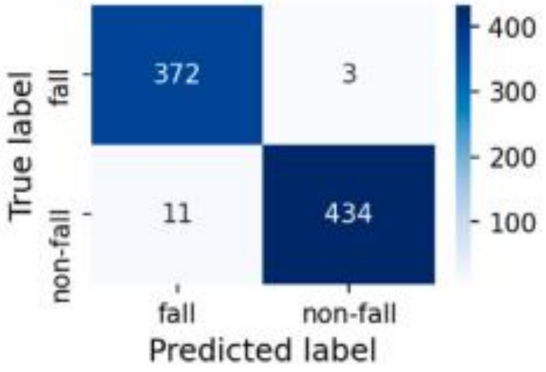
(b) TS + C2



(c) TS



(d) C1



(e) C2

4. Evaluation experiment

Fall activity recognition

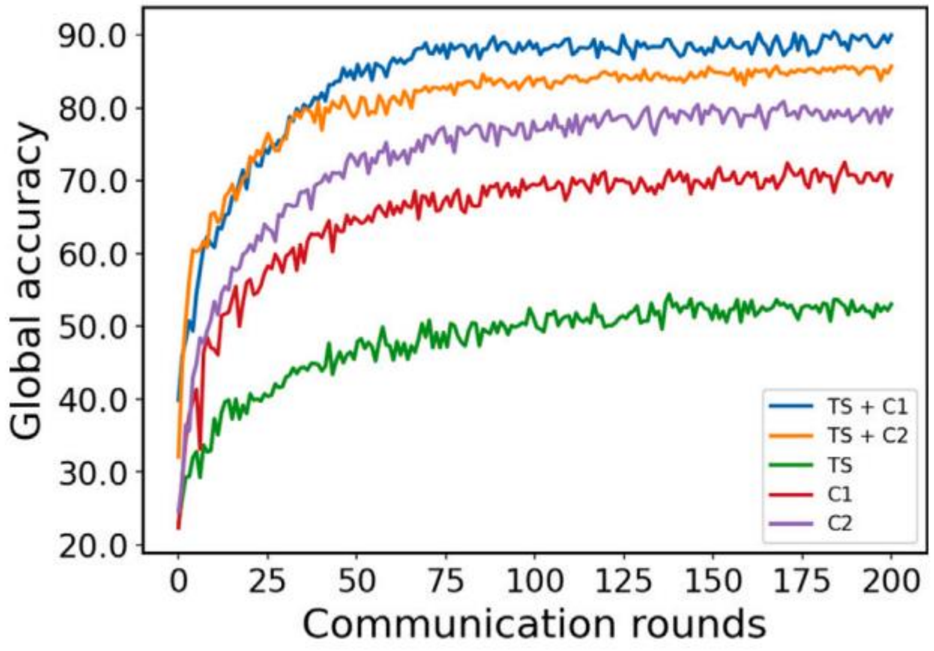


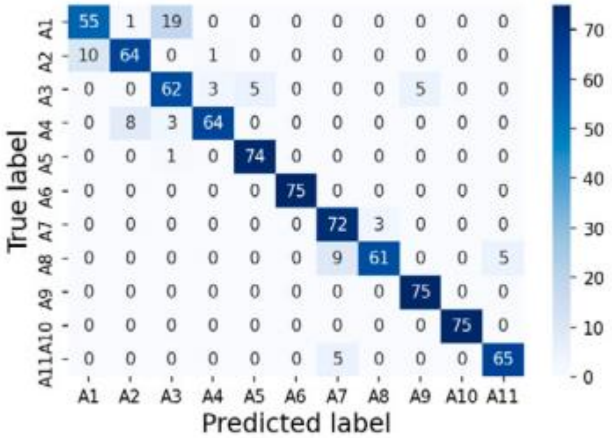
Fig. 12. The accuracy curves of global model for fall activity recognition.

Table 4
Accuracy, Precision, Recall, and F1-score for fall activity recognition.

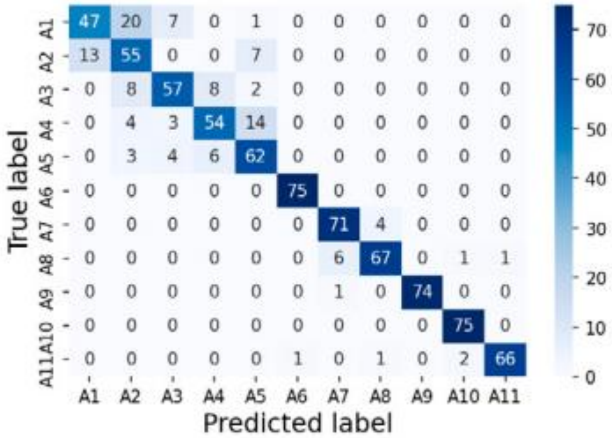
	Accuracy	Precision	Recall	F1
TS + C1	89.769 ± 0.162	90.094 ± 0.155	89.769 ± 0.162	89.748 ± 0.162
TS + C2	84.097 ± 0.234	84.546 ± 0.253	84.097 ± 0.234	83.999 ± 0.249
TS	53.085 ± 0.255	52.938 ± 0.450	53.085 ± 0.255	52.286 ± 0.360
C1	70.439 ± 0.325	70.607 ± 0.318	70.439 ± 0.325	70.175 ± 0.329
C2	77.732 ± 0.415	77.777 ± 0.410	77.732 ± 0.415	77.416 ± 0.428

4. Evaluation experiment

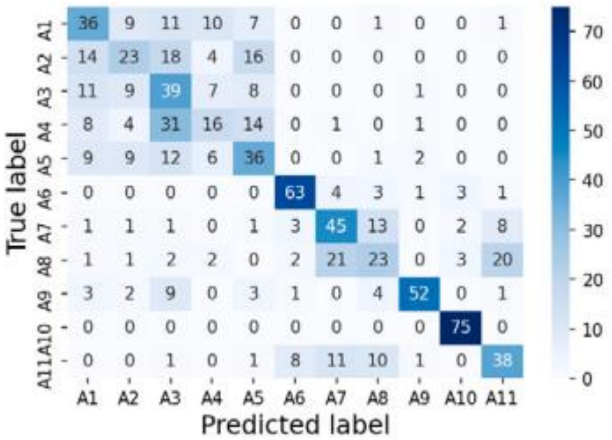
Fall activity recognition



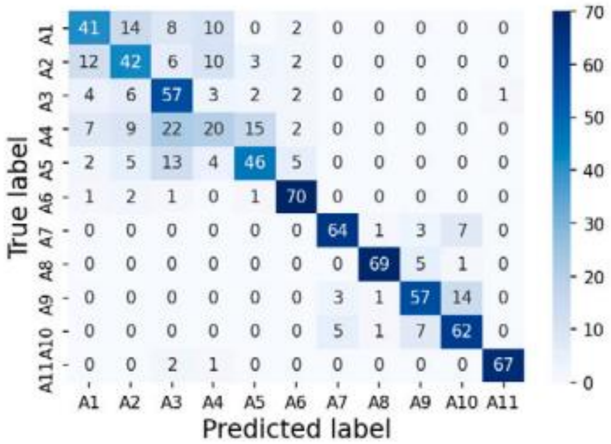
(a) TS + C1



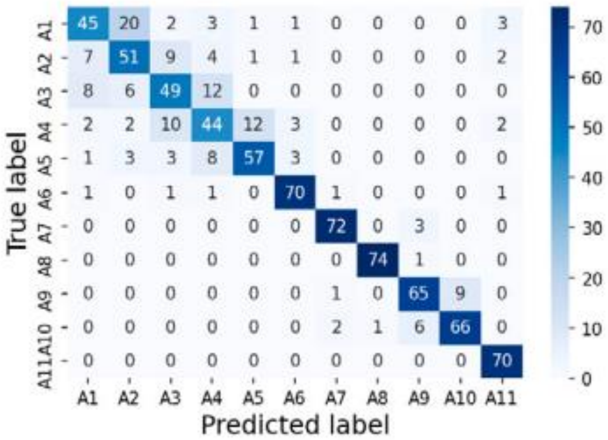
(b) TS + C2



(c) TS



(d) C1



(e) C2

4. Evaluation experiment

Discussion

Multi-modal fusion

- Input-level data fusion can provide better classification results
- If building a model with single-modal data, it's better to use image data than sensor data

Federated learning

- Protected privacy
- Low accuracy for non-IID

5 Conclusion

5. Conclusion

Conclusion

Conclusion

- Proposed FL-based multimodal data fusion method for fall detection that protects the security of user information
- Not exposing user data
- Achieves higher accuracy than single-modal data through input-level data fusion

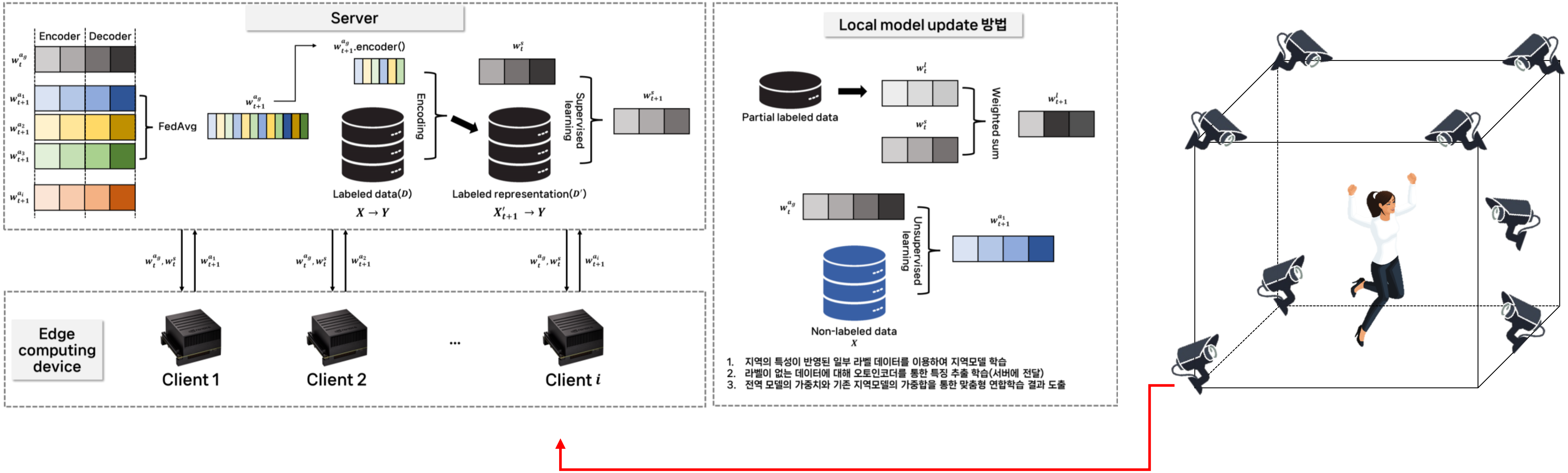
Future work

- Fusing more modalities of data
- Building Action Recognition based on federated multi-task learning for multiple purposes simultaneously
- Applied to other healthcare applications

How to apply?

How to apply?

1. Proposed an input-level data fusion method to combine one-dimensional time series data and two-dimensional visual data to achieve information complementarity
2. Proposed a FL framework that protects user privacy to ensure data security
3. Used benchmark dataset UP-Fall to measure efficiency, improving fall detection performance when compared to single-modal data



Thanks