이화여자대학교 대학원
2022학년도
석사학위 청구논문

# Diagnostic Classification Framework of Ambient Sensor Data for Aging in Place Smart Home Environments


빅 데 이 터 분 석 학 협 동 과 정
김 지 윤


2023

# Diagnostic Classification Framework of Ambient Sensor Data for Aging in Place Smart Home Environments

이 논문을 석사학위 논문으로 제출함

2022 년　12월

이화여자대학교 대학원

빅 데 이 터 분 석 학 협 동 과 정 김 지 윤

김 지 윤 의 석사학위 논문을 인준함

지도교수 강 윤 철 ____

심사위원 신 경 식 ____

민 대 기 ____

강 윤 철 ____

이화여자대학교 대학원

# Table of contents

# List of Tables

# List of Figures

# Summary

Aging in place (AIP) is being highlighted as the perception of the elderly and old age changes with the global population aging and as it becomes difficult to maintain the facility-centered policies for the elderly in the past. Smart home technology is highly likely to be used to monitor daily life movements, that is, independent daily life movements, which are essential in determining the AIP period and quality.

This study aims to develop a diagnostic classification model based on indoor ambient sensors according to the performance of the Instrumental Activities of Daily Living (IADLs) task for quantitative evaluation and prediction of the behavioral performance of the elderly. To this end, we propose a diagnostic aid function framework that enables diagnosis without expert labeling through an ensemble model using Bag-Of-Sensors features that reflect ADL determinants. In this regard, as a baseline model, a classification model using time-series sensor activation log data and ones using IADLs behavior labels from experts was set as a comparison model. As a result, using Bag-Of-Sensors features, the derivatives that reflect the IADLs assessment determinants, and the ensemble technique enhanced its function as a diagnostic framework. It was possible to achieve the predictive power equivalent to it through the Bag-Of-Sensors feature without the expert label. The proposed method could simplify the way to embed the expert's diagnostic mechanism in the diagnostic framework. Thus, through the proposed Bag-Of-Sensors feature, it was proved the possibility of constructing an efficient framework that can serve as a preliminary cognitive health diagnosis support system for AIP.

# I. Introduction

The growing perception of the elderly as independent individuals and the burdens in maintaining past facility-oriented elderly policies have led the Aging in place (AIP) to emerge as an ideal lifestyle option for the elderly. AIP is a concept that enables the elderly to spend the rest of their lives in their own home and community as long as possible. The most important thing in maintaining the quality of AIP and determining its duration is control over daily life, that is, independent activities of daily life (ADL). In general, Activities of Daily Living (ADL) or Instrumental ADL (IADL) is used to evaluate functional independence for daily living activities. Since changes in daily function and loss of control over daily life are considered precursors to severe cognitive functional problems, monitoring and recognizing the decline in IADL performance makes it possible to proactively diagnose dementia or mild cognitive impairment (MCI), not only evaluating the sustainability of one's AIP. Human Activity Recognition (HAR) technology can be seen as highly likely to be used to monitor ADL performance. For this reason, Human Activity Recognition (HAR) research on the daily life of the elderly is being actively studied in the HAR field.

Regarding the modality of HAR, video-based systems using computer vision technology and sensor-based systems are mainly used. Sensor-based methods are mainly selected in terms of efficiency rather than video-based systems that incur high costs, such as data capacity and computation. Among sensor-based systems, wearable devices and installed ambient sensors are mainly used for activity recognition.

The problem is that it is difficult to convert the collected behavioral sensor data to a quantitative value, which consists of qualitative evaluations such as 'Is the subject possible to perform an independent task?'. As it is common for Activities of Daily Living ability to be evaluated qualitatively through face-to-face observation by expert clinicians, ADL assessment is highly dependent on the expert's annotation. Difficulties in obtaining expert labels on ADL assessment connect to the Annotation Scarcity issue in HAR studies. The annotation process, which is expensive, time-consuming, and tedious, makes assigning labels to a large amount of data challenging (K. Chen et al., 2021).

Applying unsupervised and semi-supervised learning approaches has been an option to reduce the dependence on annotation. Yet, such techniques can increase disjunction between the model and experts' diagnostic mechanisms. Since activity labels contain ADL assessment mechanisms, excluding them from the training process could leave the model a black box.

Various feature extraction techniques can be conducted to complement this issue. To preserve the mechanism, the feature extraction should keep aligned with the patterns that clinicians capture when they diagnose. Thus, this paper proposes a simple frequency-based feature to replicate the assessment mechanism, with a vectorization technique borrowed from the natural language process – 'Bag-Of-Sensors'. The vectorization technique is used for feature extraction in the proposed cognitive assessment framework to handle annotation scarcity.

This paper aims to prove that the framework without assessment labels can achieve performance no less than the ones with them. To this end, this study conducts experiments to determine whether the model's predictive power can be improved by applying an expert's diagnostic mechanism using the Bag-Of-Sensors feature in the framework. Applying Bag-Of-Sensors feature captures the expert's assessment mechanism and increases the diagnostic model's domain connectivity.

This study aims to develop a diagnostic classification framework for the elderly based on ambient sensor data from the performance of daily life tasks. Most importantly, the annotation scarcity problem is supplemented through feature fusion and classifier ensemble by reflecting the diagnostic mechanism of experts. Ambient sensor data installed indoors is used for training models rather than the ones from widely used wearable sensors, especially regarding indoor activities of the elderly.

The proposed application would be a decision-supporting system to monitor and manage their independent ability to sustain their AIP lifestyle. It can suggest how much assistance in living is needed to the users or warn the user to set another following action whether to consult advanced measures or not.

# II. Related Work

## A. Algorithmic approach to ADL assessment

In general, behavioral abnormality symptom evaluation is performed with a qualitative assessment such as ‘Is it possible to perform an independent task?’ under visiting observation. Quantifying evaluation items and algorithmization of the diagnosis process are essential to embedding these experts' cognitive ability diagnosis processes in algorithms. Serna (2006) scored and analyzed KTA (Kitchen Task Assessment), a type of IADL, on six criteria for research subjects from CDR levels 0 to 3. Wojtusiak (2021) developed the Computational Barthel Index Tool (CBIT), which can automatically evaluate and predict current and future ADL scores based on the patient's medical history, to algorithmize behavioral abnormal symptom evaluation. Mlinac (2016) attempted to create a model for dementia-induced disorders and evaluate the relationship between neuropsychological function and ADL. Panhwar (2018) developed a vulnerability evaluation algorithm based on ADL picking objects from the floor to build a model to classify kinematic data on vulnerability levels in older adults. On the other hand, rather than embedding expert labels in the algorithm, such as checking whether the cluster is formed properly, it only checked whether the calculated value of the observation score and the actual diagnosis name were correlated.

## B. Dementia-related Activity Recognition

Arifoglu (2017) approached activity recognition as a sequence labeling problem and attempted to develop a model for detecting dementia-related anomalies. The anomaly detection data was generated by inserting actions that could be considered symptoms into a normal sequence. The symptoms included specific behavior repetition and negative numbers during sleep. In the case of Arifoglu (2021), the behavior was structurally approached and structured into activity and sub-activity. Abnormal symptoms were determined according to repetition, initiation, and order compliance for the activity and sub-activity, respectively. The abnormal behavior data were generated by manipulating the normal sequence according to this criterion. One of his feature representation methods was 'Bag-Of-Sensors', which resembles the bag-of-words representation method in the document recognition field. In addition to the frequency vectorization of sensor values, this paper applied a novel approach of 'Sensor Frequency – Inverse User Frequency' to featurize the rareness of each sensor value.

Both studies of Arifoglu generated features for sensor data to detect anomalous behaviors among daily behavior data, which are meaningful in approaches to conceptualize and generate activity label-related features. However, the studies were only conducted with manually generated abnormal data, of which the validity is not guaranteed, and experiment with actual anomalous data was not conducted.

Tsang (2017) attempted context-based modeling of Mild Cognitive Impairment (MCI) in an indoor behavioral tracking study using motion sensors. Although indoor behavior was classified through the distance between the wearable sensor and the position detection sensor, only changes in daily patterns were considered in detecting abnormalities through this.

## C. Smart home Ambient Sensor data analysis

Dawadi has conducted a cognitive assessment observation experiment at CASAS Smart home to collect data and has continuously researched the analysis of the diagnostic model. The study subjects were given a task of IADLs diagnostic behavior and labeled by experts. Dawadi (2011) conducted a classification model for dementia and cognitively healthy groups and concluded that both models used in the experiment showed some performance. The classification models were Naive Bayes, Decision Tree (J48), Sequential Minimal Optimization (SMO), and Neural Network.

In the case of Dawadi (2013), experiments were conducted on three groups of dementia, MCI, and Cognitively Healthy, which were more detailed than in previous studies. The quality of task performance was evaluated and learned according to the DOT score scored by experts.

The fact that he collected direct assessment observation diagnostic data and conducted models based on the collected labels is both a strength and a limit. The validity of the abnormal data, the data from actual dementia and MCI patients, is guaranteed, yet the model itself remained dependent on the activity label. That is, the model cannot predict the diagnosis of other cases without the expert's annotation. Accordingly, this study intends to use the corresponding Cognitive Assessment activity data as experimental data and build a framework that works without the expert’s annotation about the subjects’ task performance.

# III. Materials and Method

## A. Data Description

This paper used the Cognitive Assessment activity data provided by CASAS Smart Home for the framework. The data consist of ambient sensor data collected from 400 subjects who were asked to perform a given behavior task in the WSU CASAS smart apartment as part of a study conducted by Washington State University. The study intended to minimize the gap between the experimental environment and the actual assessment environment by limiting the collection of behavioral data to specific assessment tasks rather than collecting all daily living activities.

Six types of sensors are used in the analysis, including motion sensors and door opening/closing detection sensors (Table 1). Subjects visited each different date and performed under a single resident situation. A total of 24 tasks related to Activities of Daily Life were asked to be observed by the experts. Activity labels corresponding to the log were attached to each subject's task performance. Ten diagnostic labels for each subject were attached in the original study.

Among the 400 samples, 20 samples in which no data was recorded and one sample labeled with ‘diagnosis not available’ were dropped. To remove the sociodemographic factors from the analysis, this study summated the labels (labels 3~9) into the ‘Cognitively Healthy’ label, which excludes the ‘other medical’ labeled samples. The target feature is finalized into three classes – ‘Dementia,’ ‘MCI,’ and ‘Cognitively Healthy.’

| Sensor label | Description |
|---|---|
| Mxx | motion sensor |
| Ixx | item sensor for selected items in the kitchen |
| Dxx | door sensor |
| AD1-A | burner sensor |
| AD1-B | hot water sensor |
| AD1-C | other medical |

**Table 1 CASAS assessment data Sensors list**

Task 1) Sweep the kitchen and dust the living room.

1) Participant retrieves broom from supply closet
2) Participant retrieves duster from supply closet
3) Participant retrieves dust pan and brush from closet
4) Participant sweeps kitchen floor
5) Participant uses dust pan and brush
6) Participant dusts living room
7) Participant dusts dining room
8) Participant returns broom to supply closet
9) Participant returns duster to supply closet
10) Participants returns dust pan and brush to supply closet

**Figure 1 CASAS assessment task instruction example**

Figure 1 shows Task 1 out of 24 tasks as an example. Label for performing 'Participant dusts dining room,' Task 1's seventh subtask, is labeled as '1.7', as shown in Figure 2, an example from MCI diagnosed subject ID.6.

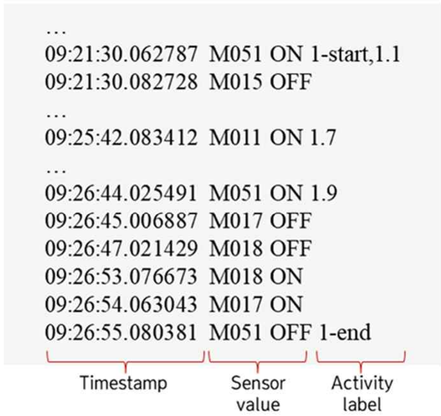



**Figure 2 CASAS assessment sensor log data example**

## B. Feature Engineering

There are six features used in experiments - Sequential sensor log (SS), Sequential time-gap log (ST), Sensor frequency (SF), Sensor frequency-inverse user frequency (SF-IUF), Sequential activity label (SA), and Activity labels feature (AF). (Table 2)

Sensor-related features correspond to the sensor activation value from the raw log data. The sequential time-gap log (ST), the time difference between the current and previous sensor activation, is to scale the raw timestamp, which varies by when each experiment was taken. The activity label-related features refer to professional clinicians' assessment mechanism.

| Abbreviation | Feature name | Description |
|---|---|---|
| SS | Sequential sensor log | Sequence (Max length 4000) |
| ST | Sequential time-gap log | Sequence (Max length 4000) |
| SF | Sensor frequency | Bag-Of-Sensors Vectorization |
| SF-IUF | Sensor frequency-inverse user frequency | Bag-Of-Sensors Vectorization |
| SA | Sequential activity label | Sequence (Max length 4000) |
| AF | Activity labels feature | Task: Initialization, Completion,<br>Sub-task: Frequency, Order |

**Table 2 Descriptions of the features used**

The dependent variable, 'target,' consists of three classes - Dementia, MCI, and Cognitively Health.

| Label # | Class name (Diagnosis) | Description |
|---|---|---|
| 1 | Dementia | 36 pax (10.8%) |
| 2 | MCI (Mild Cognitive Impairment) | 59 pax (17.8%) |
| 3 | Cognitively Healthy | 237 pax (71.4%) |

**Table 3 Descriptions of the target variable**

### 1. Sequential Sensor log

The sequential Sensor log feature corresponds to the sensor activation value. As shown in Table 2, each sequence is padded into the max length of 4000 timestamps. Since the raw data is not

# IV. Results

## A. Experimental Set-Up

The training dataset was learned by applying stratified 10-fold cross-validation according to three target classes: Dementia (36 people, 9.4%), MCI (59 people, 15.5%), and Cognitively Healthy (285 people, 75%).

Focal loss ($FL(p_t)$) is applied as training loss instead of cross-entropy loss to minimize the class imbalance issue.

$$FL(p_t) = -\alpha_t\,(1 - p_t)^{\gamma} log(p_t), \qquad \alpha = 0.25, \gamma = 2$$

The classification performance of the models is measured by weighted f1-score and accuracy. Weighted f1-score is applied as a performance measure to supplement the class imbalance issue.

## B. Evaluation of models and features

Table 5 shows the average, min, and max values of the 10-fold validation result, and the best results are presented in bold font.

### 1. Benchmarking models

Model 1 provides the baseline of the model performance level since the features it is fed with are the closest form to the raw data. Model 4-3, which shows the best result among the assumed gold standard model 'Model 4', achieves 3.3%p more average accuracy and 2.2%p more average f1 score than Model1.

vectorization method.

Similar to the BOW feature vectorization method, Bag-Of-Sensors is what all sensors are listed in the form of columns from all the samples, and each sensor's frequency value is given.

sf refers to a frequency value of a sensor within a user sample (Equation (1)), whereas uf is a value that indicates how common a sensor is across the set of user samples (Equation (2)). Inverse-uf (iuf) can be obtained by dividing the total number of users by the number of user samples that include the corresponding sensor and then taking the logarithm (Equation (3)). sf-iuf value is obtained by multiplying sf and iuf (Equation (4)). 'The higher the frequency of a sensor in a specific user sample' and 'the smaller the number of samples including that sensor among all user samples,' the higher the sf-iuf value. Therefore, using this value makes it possible to obtain the effect of filtering out common sensors in all samples.

$$sf(sensor, user) = frequency(sensor, user) \tag{1}$$

$$uf(sensor) = |\{user \in User : sensor \in user\}| \tag{2}$$

$$iuf(user, sensor) = log\frac{User}{1+uf(sensor)} \tag{3}$$

$$sf - iuf(sensor, User) = sf(sensor, user) \cdot iuf(sensor, User) \tag{4}$$

As each sensor activation value is a unique non-numeric value in the raw sensor data, in the form of a character, the Bag-Of-Sensors concept can be applied by simply using the BOW module used in the text-mining process. CountVectorizer and TfidfVectorizer from the scikit-learn package were used to vectorize each sensor value.

## 3. Activity Label data featurization

CASAS data provides an activity label sequence attached to the sensor log sequence. Defining it as the 'Sequential Activity label (AS),' its maximal length is set to 4000 timesteps, like other sequential features, padded with 0. The activity labels feature (AF) is derivate from the sequential activity label data by the criteria of Serna(2006) to reflect how the labels are attached through the ADL assessment mechanism. Following the criteria of how Serna(2006) scored the subject's performance and how Arifoglu(2021) approached an activity as a vertical structure, AF is built upon the criteria of Initialization and Completion of the tasks, and Frequency, Order of the sub-task. How the feature is constructed and vectorized is shown in Figure 4.

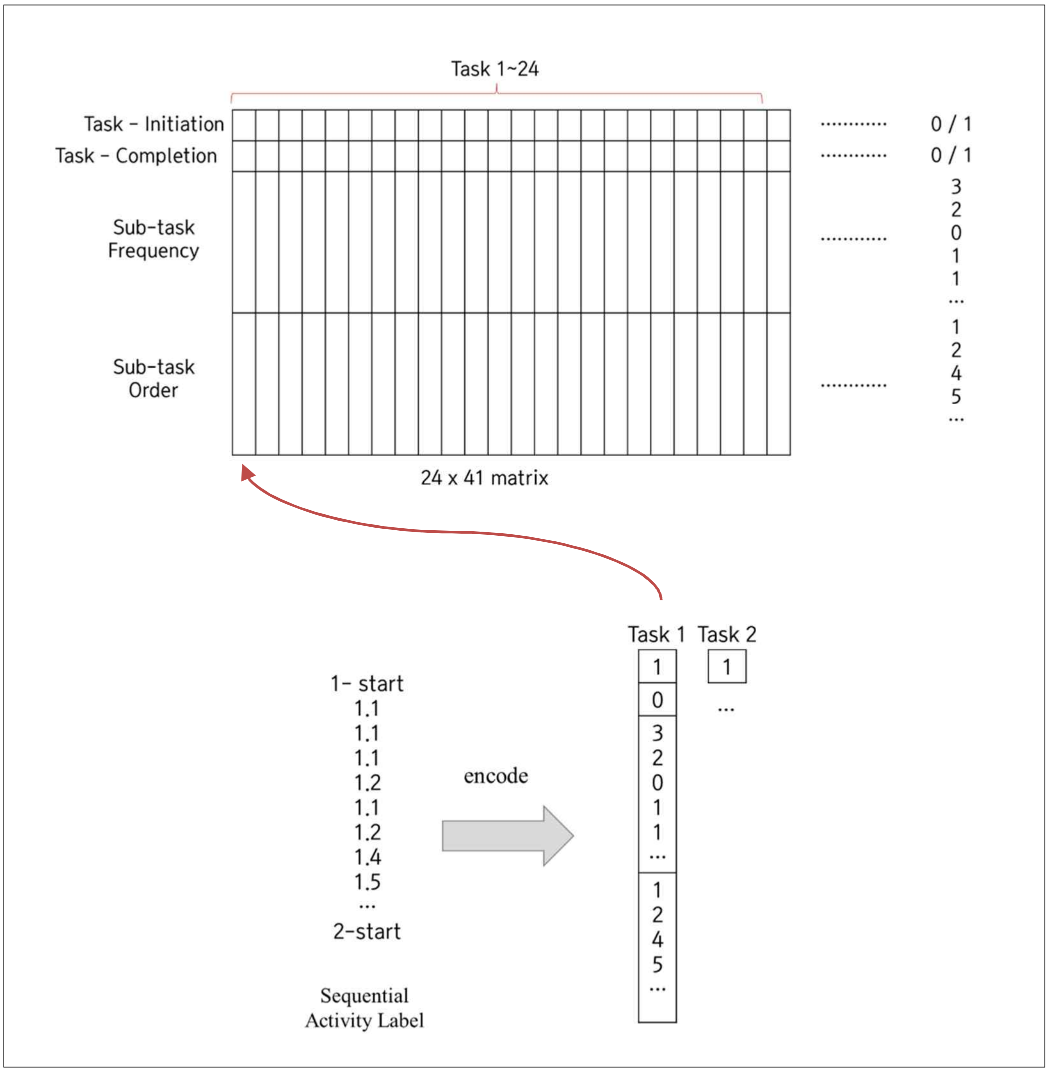



**Figure 4 Formation of Activity labels feature**

Applying the AF feature contrasts with Bag-Of-Sensors features. It investigates whether the independence of activities in the daily life of the elderly can be assessed without an expert's observation.

## C. Diagnostic Classification Framework

This paper applied LSTM (Long Short Term Memory) and CNN (Convolutional Neural Network) layers to construct the experimental diagnostic classification models. LSTM refers to the structure of a neural network that enables long-term/short-term memory with feedback connection, unlike conventional vanilla RNN which suffers from vanishing long-term memory. CNN is a type of Artificial Neural Network that uses convolutional operations, which enables maintaining the spatial information within multidimensional data to the next layer. Convolution is an operation to extract a feature within multidimensional data such as an image, which allows learning directly from the data without having to manually extract features.

In this paper, LSTM and CNN were applied to learn the sequential features (SS, ST, SA) and the generated features (SF, SF-IUF, AF) respectively. Learning the generated features with CNN is to fully utilize its advantages- capturing the pattern from the part rather than the whole matrix, and from the relationship between one element and the surrounding elements.

### 1. Benchmarking

The purpose of this study is to prove whether Bag-Of-Sensors vectorized feature can improve domain connectivity and performance of the diagnostic framework. To compare, a model only with sensor sequence, the rawest state of the collected data, will be the baseline model. This baseline model, Model 1, is constructed with LSTM layers to directly classify the subject's state from the sequential sensor data.

To set the gold standard comparator of the suggested models, models of Model 4 are fed with expert label data. These are without the Bag-Of-Sensors features to investigate the performance difference by the presence of it. To classify the subject's state from the Activity Label feature, CNN layers are applied, along with LSTM layers as an ensemble classifier when the model uses the sequential features (SS, ST, SA) as well. Details of benchmarking models can be seen in Figure 5 as Model 1 and Model 4.

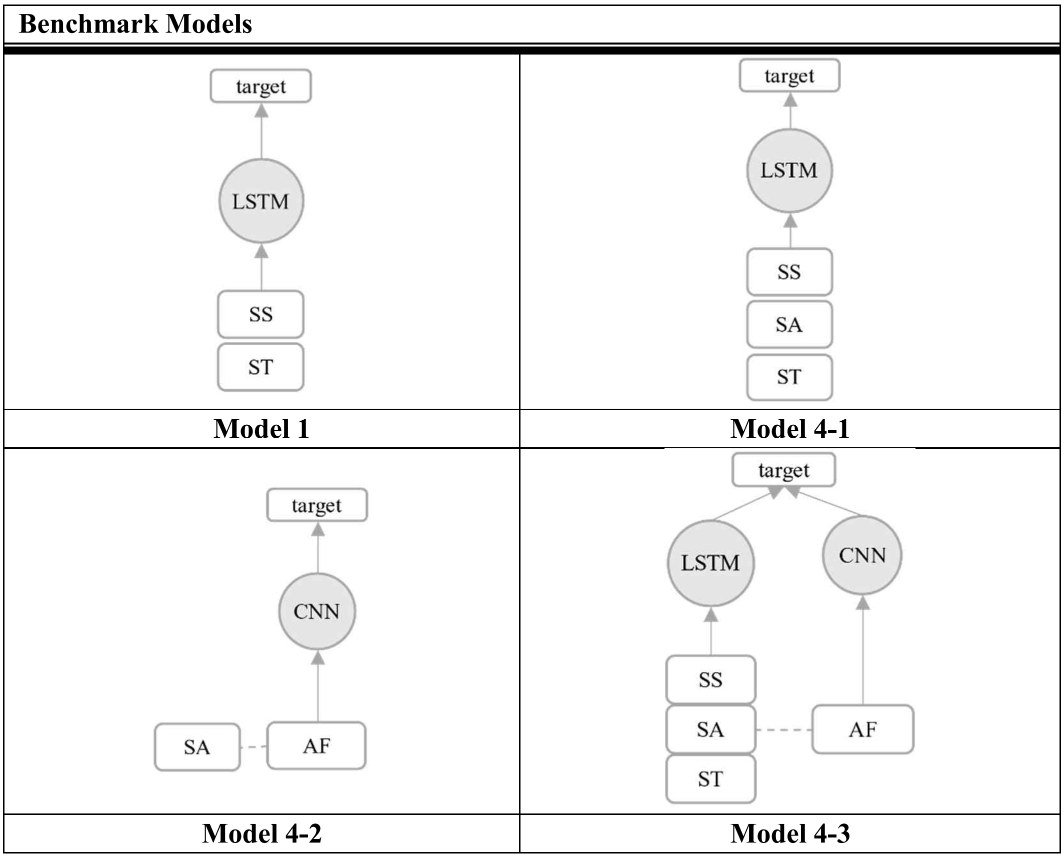



**Figure 5 Benchmarking models**

## 2. Bag-Of-Sensors Classifiers

The proposed method named ‘Bag-Of-Sensors Classifiers’ uses the Bag-Of-Sensors features, SF and SF-IUF, to prove that its role as the key factor can improve domain connectivity and performance of the diagnostic framework.

Model 2 uses only the Bag-Of-Sensors features, experimenting by using them together and separately. Model 3 is an ensembled version of model 2l2 with the SS and ST, experimenting by using SF and SF-IUF together and individually. Details of the proposed method can be seen in Figure 5 as Model 2 and Model 3.

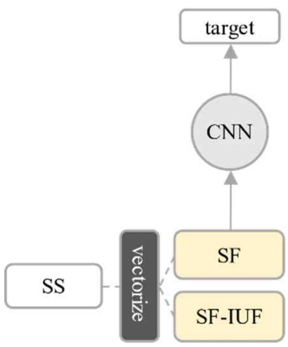
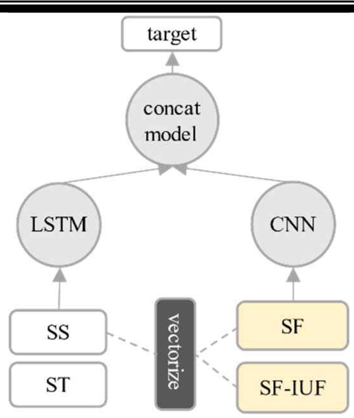
| **Bag-Of-Sensors Classifiers** | |
|---|---|
|  |  |
| **Model 2** | **Model 3** |

**Figure 6 Bag-Of-Sensors Classifiers**

Descriptions of the features and the model’s structure are summarized in Table 4.

| Model # | | Structure | Features |
|---|---|---|---|
| Model 1 | 1-1 | LSTM | Sequential Sensor log (SS)<br>Sequential time-gap log (ST) |
| Model 2 | 2-1 | CNN | Sensor frequency (SF)<br>Sensor frequency-inverse user frequency (SF-IUF) |
| | 2-2 | CNN | Sensor frequency (SF) |
| | 2-3 | CNN | Sensor frequency-inverse user frequency (SF-IUF) |
| Model 3 | 3-1 | LSTM + CNN | Sequential Sensor log (SS)<br>Sequential time-gap log (ST)<br>Sensor frequency (SF)<br>Sensor frequency-inverse user frequency (SF-IUF) |
| | 3-2 | LSTM + CNN | Sequential Sensor log (SS)<br>Sequential time-gap log (ST)<br>Sensor frequency (SF) |
| | 3-3 | LSTM + CNN | Sequential Sensor log (SS)<br>Sequential time-gap log (ST)<br>Sensor frequency-inverse user frequency (SF-IUF) |
| Model 4 | 4-1 | LSTM | Sequential Sensor log (SS)<br>Sequential time-gap log (ST)<br>Sequential activity label (SA) |
| | 4-2 | CNN | Activity labels feature (AF) |
| | 4-3 | LSTM + CNN | Sequential Sensor log (SS)<br>Sequential time-gap log (ST)<br>Sequential activity label (SA)<br>Activity labels feature (AF) |

**Table 4 Description of the Models used**

# IV. Results

## A. Experimental Set-Up

The training dataset was learned by applying stratified 10-fold cross-validation according to three target classes: Dementia (36 people, 9.4%), MCI (59 people, 15.5%), and Cognitively Healthy (285 people, 75%).

Focal loss ($FL(p_t)$) is applied as training loss instead of cross-entropy loss to minimize the class imbalance issue.

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} log(p_t), \qquad \alpha = 0.25, \gamma = 2$$

The classification performance of the models is measured by weighted f1-score and accuracy. Weighted f1-score is applied as a performance measure to supplement the class imbalance issue.

## B. Evaluation of models and features

Table 5 shows the average, min, and max values of the 10-fold validation result, and the best results are presented in bold font.

### 1. Benchmarking models

Model 1 provides the baseline of the model performance level since the features it is fed with are the closest form to the raw data. Model 4-3, which shows the best result among the assumed gold standard model 'Model 4', achieves 3.3%p more average accuracy and 2.2%p more average f1 score than Model1.

## 2. Bag-Of-Sensors Classifier

Comparing the baseline Model 1 and proposed methods Models 2 and 3, the performance improvement was achieved by applying BOS features. Results from Model 3 show that BOS features can also improve the model performance through an ensemble classifier. It was found that the performance of Model 2, which did not use sequential features, was efficiently improved while reducing their training time significantly. Model 2-1 and Model 2-3 even achieved better performance than Model 3.

Analyzing the detailed model results of each Model 2 and 3, results were better when SF and SF-IUF were fed to the model together. In the case of Model 2, Model 2-1 showed the best average accuracy with a smaller range of values than model 2-3. It showed 0.2%p less f1 score, yet 8%p less range of values than Model 3-2, which makes it the best among Model 2. In the case of Model 3, Model 2-1 showed the best average accuracy and f1 score.

Among the proposed Model 2 and Model 3, some models performed at the level of Model 4, which requires expert label diagnosis. Especially Model 2-1 and Model 2-3 meet the performance level of Model 4-3, which is the gold standard model.

Overall results suggest that the expert diagnosis mechanism could be efficiently embedded with the BOS features. Figure 8 summarizes the performance differences between the bests of each Model group.

| Model | Feature | Accuracy | | | F1 score | | | Training time (sec) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | average | min | max | average | min | max | average | min | max |
| Model 1 | SS, ST | 0.741 | 0.576 | 0.788 | 0.655 | 0.554 | 0.694 | 1709.568 | 1015.105 | 2119.791 |
| Model 2-1 | SF, SF-IUF | **0.771** | 0.727 | 0.818 | **0.677** | 0.638 | 0.757 | 18.374 | 7.238 | 35.800 |
| Model 2-2 | SF | 0.762 | 0.697 | 0.848 | 0.667 | 0.573 | 0.779 | 27.500 | 10.034 | 37.985 |
| Model 2-3 | SF-IUF | 0.771 | 0.697 | 0.848 | 0.679 | 0.573 | 0.779 | 26.820 | 9.915 | 48.042 |
| Model 3-1 | SS, ST, SF, SF-IUF | 0.768 | 0.697 | 0.848 | 0.675 | 0.573 | 0.779 | 1484.609 | 598.182 | 2078.228 |
| Model 3-2 | SS, ST, SF | 0.768 | 0.697 | 0.848 | 0.670 | 0.573 | 0.779 | 1445.937 | 837.883 | 2456.677 |
| Model 3-3 | SS, ST, SF-IUF | 0.762 | 0.697 | 0.848 | 0.669 | 0.573 | 0.779 | 1143.618 | 828.132 | 1636.514 |
| Model 4-1 | SS, ST, SA | 0.750 | 0.667 | 0.848 | 0.667 | 0.573 | 0.779 | 2287.898 | 1610.998 | 3779.002 |
| Model 4-2 | AF | **0.774** | 0.697 | 0.848 | **0.678** | 0.573 | 0.779 | 36.790 | 26.150 | 50.301 |
| Model 4-3 | SS, ST, SA, AF | 0.773 | 0.758 | 0.788 | 0.674 | 0.653 | 0.694 | 2257.695 | 1269.170 | 3160.747 |

**Table 5 Performance Evaluations of models**

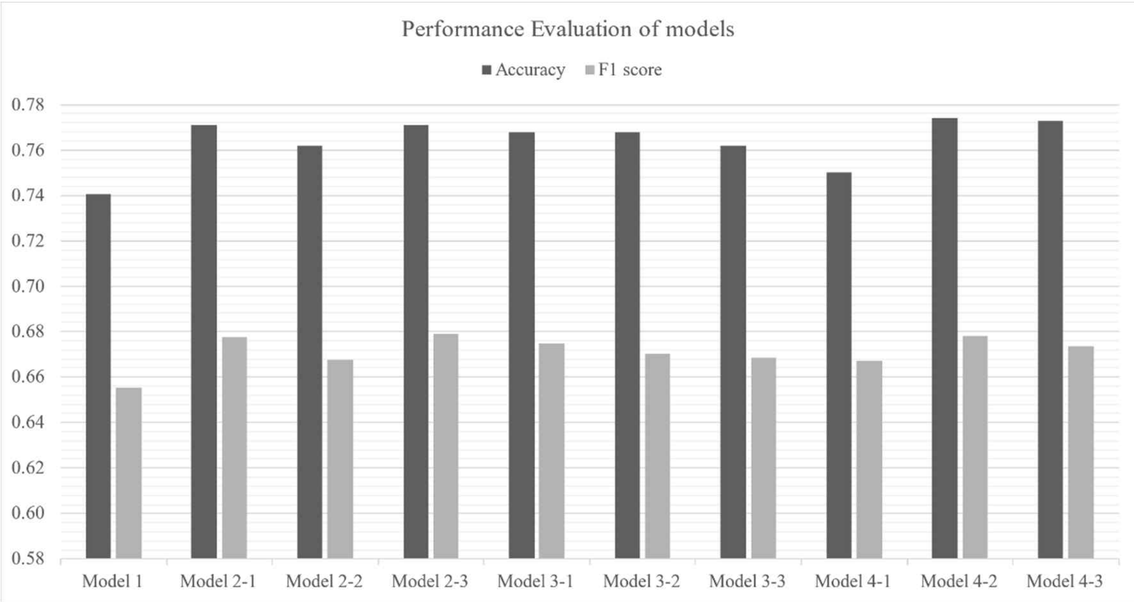



**Figure 7 Performance Evaluation of models**

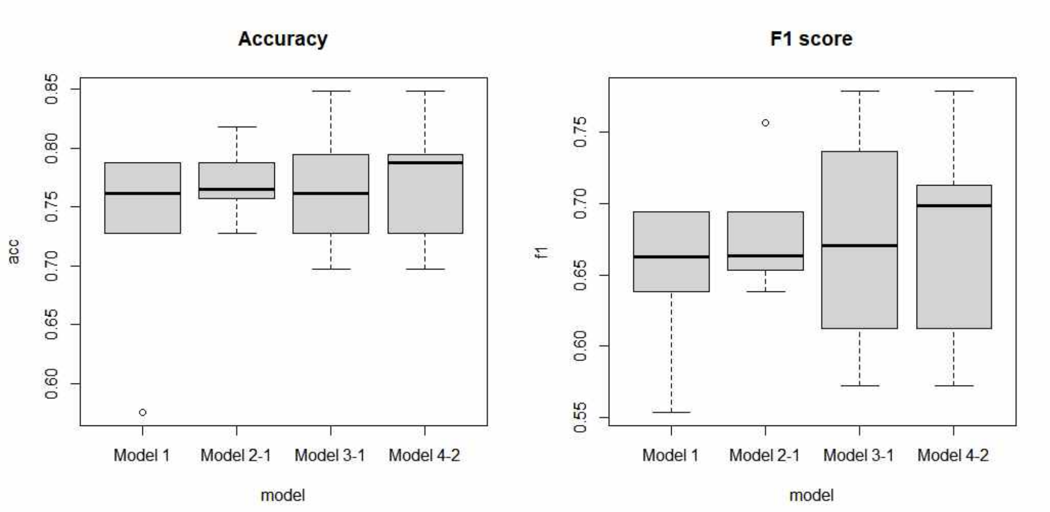



**Figure 8 Performance of the bests of each Model group**

## 3. Performance Improvement Evaluation

Table 6 and Table 7 show the result of the T-Test applied to evaluate the performance improvements of each suggested model (Model 2-1, Model 3-1). The results show that the improvements in the performance that the suggested models made were not significant enough. The result of ANOVA applied by models also shows that performance differences between the models are not significant enough.

| | *df* | *t* | *p* | 95% CI |
|---|---|---|---|---|
| Accuracy | 18 | -1.412 | 0.175 | -0.075, 0.015 |
| F1 score | 18 | -1.286 | 0.215 | -0.058, 0.014 |

**Table 6 T-Test between Model 1 and Model 2-1**

| | *df* | *t* | *p* | 95% CI |
|---|---|---|---|---|
| Accuracy | 18 | -1.104 | 0.284 | -0.079, 0.025 |
| F1 score | 18 | -0.77733 | 0.447 | -0.071, 0.033 |

**Table 7 T-Test between Model 1 and Model 3-1**

| | *df* | Sum of Squares | Mean Square | *F* | *p* |
|---|---|---|---|---|---|
| Accuracy | | | | | |
| model | 3 | 0.007 | 0.002 | 1.062 | 0.377 |
| Residuals | 36 | 0.079 | 0.002 | | |
| F1 score | | | | | |
| model | 3 | 0.003 | 0.001 | 0.403 | 0.752 |
| Residuals | 36 | 0.102 | 0.003 | | |

**Table 8 ANOVA by model (Model 1, 2-1, 3-1, 4-2)**

# V. Discussion

Although proposed models with BOS features outperform the benchmarking models, one disadvantage of our model is that they are built with limited data. Since it is a chronic problem in the HAR field, models were evaluated with 10-fold cross-validation. Yet, the data drift problem remains according to measuring environment, subject, and object. Retaining sufficient training data in further study may allow the models to reinforce their validity and performance.

Another disadvantage of our proposed method is that they lack explainability. Since the models are based on deep neural network, it remains a black box to the users and cannot provide further information about the model's result or variable importance. Applying SHAP Gradient Explainer or LIME algorithm may complement the framework as an explainable model and strengthen its function as a decision support system.

# VI. Conclusions

In accordance with the rapidly aging population, the lives of the elderly have changed significantly. Meanwhile, AIP has risen as a beneficial option that has many advantages. The elder and society prefer to carry on their lives within their own place and community rather than spending the rest of their lives in facilities.

Growing attention to AIP leads to a housing environment for the elder lifestyle. For the elderly to lead their lives, their ability to live independently and a system to support them are crucial. It would be most desirable if everyone could access sufficient assisted living services, but such resources are limited. In this situation, smart home system can help with resource allocation.

This paper proposes the framework as a decision-supportive system by giving preliminary ADL results. With prominent classification performance, efficiency is an essential value for the preliminary diagnostic model. Experiments to investigate whether the expert's diagnosis mechanism can be embedded in the model through Bag-Of-Sensors features showed that Bag-Of-Sensors features can improve the classification performance, also reducing its training time. Through using Bag-Of-Sensors features, the derivatives that reflect the IADLs assessment determinants, and the ensemble technique, the framework enhanced its function as a diagnostic framework. This proposed feature could simplify the way to embed the expert's diagnostic mechanism in the diagnostic framework.

# Bibliography


Arifoglu, Damla; BOUCHACHIA, Abdelhamid. 2017. “Activity recognition and abnormal behaviour detection with recurrent neural networks.” *Procedia Computer Science.* 110: 86-93.

Arifoglu D, Wang Y, Bouchachia A. 2021. “Detection of Dementia-Related Abnormal Behaviour Using Recursive Auto-Encoders.” *Sensors (Basel)*.

Audrey Serna, Vincent Rialle, Hélène Pigot. 2006. “Computational representation of Alzheimer's disease evolution applied to a cooking activity.” *Studies in Health Technology and Informatics*.

Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z., & Liu, Y. 2021. “Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities.” *ACM Computing Surveys (CSUR)*. 54(4), 1-40.

Damla Arifoglu, Abdelhamid Bouchachia. 2017. “Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks.” *Procedia Computer Science.*

Dawadi, P., Cook, D., and Schmitter-Edgecombe, M. 2013. “Automated cognitive health assessment using smart home monitoring of complex tasks.” *IEEE Transactions on Human-Machine Systems*.

Dawadi, P., Cook, D., Parsey, C., Schmitter-Edgecombe, M., and Schneider, M. 2011.

"An approach to cognitive assessment in smart homes." *KDD workshop on Medicine and Healthcare.*

D. Cook, A. Crandall, B. Thomas, and N. Krishnan. 2013. "CASAS: A smart home in a box," *IEEE Computer.*

Kee, B. S., Youn, Y. C., Kim, Y. I., & Park, S. Y. 2015. "The prediction model for behavioral and psychological symptoms of dementia in Alzheimer's disease: Preliminary study." *INTERNATIONAL PSYCHOGERIATRICS*, 27, S137-S137.

Mlinac ME, Feng MC. 2016. "Assessment of Activities of Daily Living, Self-Care, and Independence." *Arch Clin Neuropsychol*. 31(6):506-516.

Seelye, A., Schmitter-Edgecombe, M., Cook, D., and Crandall, A. 2013. "Naturalistic assessment of everyday activities and smart environment prompting technologies in mild cognitive impairment subtypes." *Journal of the International Neuropsychological Society*

Serna, A., Pigot, H., & Rialle, V. 2007. "Modeling the progression of Alzheimer's disease for cognitive assistance in smart homes." *User Modeling and User-Adapted Interaction*, 17(4), 415-438.

Urwyler, P., Stucki, R., Rampa, L., Müri, R., Mosimann, U. P., & Nef, T. 2017. "Cognitive impairment categorized in community-dwelling older adults with and without dementia using in-home sensors that recognise activities of daily living." *Scientific reports*, 7(1), 1-9.

Wojtusiak J, Asadzadehzanjani N, Levy C, Alemi F, Williams AE. 2021. "Computational Barthel Index: an automated tool for assessing and predicting activities of daily living among nursing home patients." *BMC Med Inform Decis Mak.* 21(1):17. Published 2021 Jan 9.

Y. N. Panhwar, F. Naghdy, D. Stirling, G. Naghdy, and J. Potter. 2018. "Quantitative Frailty Assessment Using Activity of Daily Living (ADL)." *2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE)*. pp. 269-272.

Yang, Y., Yang, H. D., Hong, Y. J., Kim, J. E., Park, M. H., Na, H. R., ... & Kim, S. 2012. "Activities of daily living and dementia." *Dementia and Neurocognitive Disorders*, 11(2), 29-37.

# 국문초록

김지윤
빅데이터분석학협동과정
이화여자대학교 대학원

전 세계적인 고령화에 따라 노인과 노년기에 대한 인식이 변화하고 과거 시설 중심의 노인정책 유지가 어려워지면서, '노인의 지역사회 계속 거주 (Aging in place, 이하 AIP)'가 부각되고 있다. AIP 기간과 그 질을 결정하는 데에 중요한 일상에 대한 통제력, 즉 독립적 일상생활 동작을 모니터링하는 데에 스마트홈 기반 Human Activity Recognition 기술은 그 활용 가능성이 높다.

본 연구는 노인의 행동 수행 능력에 대한 정량적 평가 및 예측을 위해, 일상생활수행능력(Instrumental Activities of Daily Living, 이하 IADLs) task 수행에 따른 실내 ambient sensor(이하 주변센서) 기반 진단 분류 모델을 개발하고자 한다. 이를 위해, 센서 활성화 시퀀스와 IADL 결정요인을 반영한 Bag-Of-Sensors feature를 사용한 앙상블 모델을 통해 전문가의 라벨링 없이도 진단이 가능한 진단 보조 기능의 프레임워크를 제안한다. 이에 대해 베이스라인 모델로서 개인별 IADLs 행동 수행에 대한 시계열 센서 활성화 로그 데이터와, 이에 대한 전문가의 IADLs 행동 라벨까지 활용한 분류 모델을 비교 모델로 설정하고, Bag-Of-Sensors feature 사용 여부에 따른 성능을 평가하였다. 그 결과, 전문가 라벨을 학습 데이터로 활용하지 않더라도 Bag-Of-Sensors feature를 통해 그에 준하는 예측력을 확인할 수 있었으며, 전문가 진단 매커니즘 내재를 단순화 할 수 있었다. 이로써, 제안된 Bag-Of-Sensors feature를 통해 AIP를 위한 인지 건강 예비 진단 및 의사결정 보조 시스템 역할을 할 수 있는 효율적 프레임워크 구축의 가능성을 확인할 수 있었다.