symmetry

MDPI

Article

SAS-Net: An Agitated Behavior Early Warning Model for Community-Dwelling Dementia Patients Based on Symmetric Autoencoders and Spatio-Temporal Network

Jing Xu [1,2,*], Bin Li [3], Ping Feng [1,2], Yonghan Zhang [4] and Shengchun Yang [3]

1 Key Laboratory of Intelligent Rehabilitation and Barrier-Free for the Disabled (Changchun University), Ministry of Education, Changchun 130022, China
2 College of Computer Science and Technology, Changchun University, Changchun 130022, China
3 School of Computer Science, Northeast Electric Power University, Jilin 132011, China
4 School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK
* Correspondence: xuj81@ccu.edu.cn

Abstract

Using home sensors to provide agitation warnings for community-dwelling dementia patients without ongoing clinical supervision can enable their carers to intervene early during the agitation latent period, thereby reducing unnecessary hospital admissions and harmful events. Most existing studies are based on behavioral, sleep, and physiological data from the previous 24 h to predict patients' agitation events, which fail to fully capture patients' recent behavioral details. In this study, monitoring data from the previous 8×24 h for dementia patients are used to achieve in-depth mining of patients' living habits. Furthermore, to address the common clinical problem of extreme imbalance between agitation and normal samples (agitation samples often are extremely scarce), we designed a two-stage agitation early warning model based on symmetric autoencoders and a spatio-temporal network, dubbed SAS-Net. In the first stage, we randomly sample 80% of normal samples and employ multiple symmetric autoencoders to perform feature transformation and pseudo-label construction, and then pre-train the spatio-temporal learning network composed of convolutional networks and a gated Multilayer Perceptron. It aims to learn the patient data's intrinsic structure, underlying patterns, and effective feature extraction methods. In the second stage, we freeze the parameters of the spatio-temporal learning network and use a balanced dataset consisting of the remaining 20% of normal samples and all agitation samples to reconstruct and fine-tune the top fully connected classifier to improve the recognition performance of agitation samples. The two-stage strategy resolves the problem of ineffective training faced by deep learning models on imbalanced datasets. The experimental results demonstrate the effectiveness of the proposed SAS-Net for agitated behavior early warning.

Keywords: community-dwelling; dementia patients; patients with Alzheimer disease; agitation behavior; early warning

Check for updates

Academic Editor: George Karabatsos

Received: 25 May 2026
Revised: 21 June 2026
Accepted: 25 June 2026
Published: 29 June 2026

1. Introduction

As global aging accelerates, Alzheimer's disease and dementias have become a major public health challenge. According to statistics, the number of people with dementia worldwide is steadily rising, will peak over the next decade [1], and will place a heavy burden on healthcare systems. Therefore, there is an urgent need for targeted solutions to alleviate this growing crisis.

Dementia patients often show symptoms like memory decline, cognitive impairment, emotional abnormalities, and agitation. Among these, agitation is especially common and harmful. When agitated, patients may become restless, easily angered, talk harshly, act physically aggressive, or even hurt themselves or others. This seriously affects their safety and quality of life [2,3].

Studies show that more than 60% of dementia patients worldwide choose to stay at home for care [4,5]. However, in community-dwelling settings, there is often no professional doctor to watch over them all the time. Their mood changes and unusual behaviors may not be noticed in time, and their agitation events can be hard to predict or manage effectively. This poses a serious threat to their safety and property. This paper aims to build an effective way to detect agitation using home sensor data, in order to lower the safety risks for patients and reduce the burden on their family.

In recent years, a host of researchers have tried to employ machine learning techniques for agitation warning [6–8]. However, these techniques have many limitations. They are not efficient at handling high-dimensional, large-scale data, and they are poor at discovering the hidden relationships behind multi-source data. In contrast, deep learning, with its multi-layer network structure and strong ability to fit non-linear features, shows greater promise in agitation warning research. Nevertheless, existing studies still face two key problems: insufficient research on agitation warning based on long-term time-series data and extreme imbalance in clinical data.

To address the above problems, we used previous 8×24 h patient data and built a two-stage agitation warning network named SAS-Net based on symmetric autoencoders and a spatio-temporal learning network. This study disagrees that “agitation is only related to short-term sleep quality, physiological, and behavioral features,” because this notion often ignores the long-term causes of agitation. A large number of medical studies [9–12] have shown that agitation is essentially an impulsive behavior formed over a long period and has a time accumulation effect. Therefore, this study uses physiological, sleep, and behavioral data collected over 8×24 h (8 days) to predict agitation on the next day. Second, the clinical agitation samples are extremely sparse, resulting in the dataset facing a severe class imbalance problem, which greatly increases the learning difficulty for the warning network. To solve this, we trained the proposed SAS-Net in two stages. First, we used multiple symmetric autoencoders in SAS-Net to augment normal samples, and then pre-trained the machine learning model on the augmented data. After that, we fine-tuned the machine learning model using the remaining balanced samples, to prevent the performance of the deep learning models from collapsing when used with imbalanced datasets.

At the same time, to better understand multi-source time-series data, we designed a spatio-temporal learning network (STLN) for the proposed SAS-Net. In the time dimension, STLN uses one-dimensional convolutional neural networks (1DCNNs [13]) to extract temporal features. In the space dimension, STLN uses a gated Multilayer Perceptron (gMLP [14]) to extract spatial interaction features from multi-source data, to more fully capture early signs of agitation.

The contributions of this paper are summarized as follows:

(1) We propose using continuous physiological, sleep, and behavioral data collected over 8×24 h (i.e., 8 days) for next-day agitation warning research, and perform time alignment, splicing, and multi-source data reconstruction.
(2) We propose a two-stage deep learning model, dubbed SAS-Net, which integrates symmetric autoencoders and a spatio-temporal learning network, to alleviate the difficulty of learning under extremely imbalanced samples.

(3) We build a spatio-temporal learning network (STLN) to fully extract long-term temporal features from continuous 8-day data. Through joint interactive modeling of physiological, sleep, and behavioral data, STLN achieves a more reliable method for agitation event warnings.

(4) We conduct extensive experiments on the publicly available TIHM [15] dataset, and the experimental results verify the advancement and rationality of our model.

The rest of this paper is structured as follows: Section 2 reviews related work on agitation warning for community-dwelling dementia patients. Section 3 describes the proposed SAS-Net model in detail. Section 4 verifies the effectiveness and rationality of the SAS-Net method through extensive experiments. Section 5 discusses the optimal setting for the number of autoencoders. Finally, Section 6 provides a summary of this paper.

2. Related Works

Agitation warning technology is significant for community-dwelling dementia patient care. In recent years, researchers have conducted many studies on agitation warnings using multi-dimensional data such as behavior, sleep, and physiology data [16,17]. In the field of shallow computing and machine learning, Fletcher-Lloyd [18] et al. proposed a sensor-based agitation warning method for dementia patients living at home. By collecting environmental and physiological multi-dimensional data, they built single-label and multi-label classification models to achieve effective agitation risk warnings. Umah [19] et al. proposed an agitation warning model for dementia patients that integrates IoT and AI. They used an IoT platform to collect home sensor data such as PIR sensors and sleep monitoring pads, and uploaded the data to the cloud for processing. The model first performed data fusion, feature extraction, and feature selection on the dataset, and then built a weighted ensemble voting warning model that combines many classifiers such as K-nearest neighbors and support vector machines, etc. Hasan [20] et al. proposed a real-time positioning-based agitation warning model for dementia patients. They collected two-dimensional location data of patients using ultra-wideband positioning wristbands and combined trajectory features with clinical indicators, and built a random forest classifier, which distinguished clinically significant motor agitation from normal motor behavior in 8 h cycles, achieving an area under the ROC curve of 0.81. Khan [21] et al. built a new multi-modal sensor dataset and a baseline model for agitation warning in dementia patients. They used Empatica E4 wristbands to collect multi-modal physiological data including acceleration, blood volume pulse, electrodermal activity, and skin temperature, combined with nursing records and video annotations to complete label calibration at 1 min granularity. The system used Butterworth low-pass filtering for noise reduction and uniform resampling to 64 Hz for preprocessing. After extracting time-domain and frequency-domain features, they trained a cost-sensitive random forest classifier, and finally fused agitation and normal data to achieve a successful warning performance. Ramesh et al. [22] proposed an agitation warning method for dementia patients based on sleep and physiological data. Using key features such as sleep heart rate, respiratory variability, and time spent in bed, they used LOCF to fill missing values and optimized model training through stratified sampling and five-fold cross-validation. This method compared several ensemble models, selected ensemble learning methods such as LightGBM to build the classifier, and used SHAP values to achieve model interpretability. Finally, experiments showed that the model could effectively use sleep and physiological indicators from the previous day to warn about next-day agitation risk.

With the development of artificial intelligence, deep learning techniques have gradually been employed for agitation warnings [23,24]. Rezvani et al. [25] proposed a semi-supervised learning model to identify agitation risk in dementia patients. By combining self-supervised learning with ensemble classification, they addressed the problems of insufficient labeling and poor generalization in home sensor data. Pragadeeswaran et al. [26] proposed an IoT- and artificial neural network-based agitation warning system for dementia patients. They used IoT sensors to continuously collect environmental and physiological data such as room temperature, noise, heart rate, and activity patterns. After signal preprocessing, the data were fed into a customized ANN model to recognize agitation states and provide early warnings, thereby supporting early intervention by caregivers, reducing the frequency and severity of agitation episodes, and improving care quality and resource utilization for dementia patients. HekmatiAthar et al. [27] proposed a data-driven agitation warning method for dementia patients. They collected five-dimensional home environmental time-series data including light, noise, temperature, humidity, and air pressure, used a custom interpolation method to fill missing values, applied downsampling to handle class imbalance, and built LSTM and MLP models combined with PCA for agitation warning. Badawi et al. [28] proposed a multi-modal model for agitation and aggressive behavior warning in dementia patients. They used EmbracePlus wristbands to collect physiological signals and privacy-preserving videos to extract skeleton key points, and combined Extra Trees with GRU deep learning models to achieve real-time warning. This model can capture early signs of agitation, and balance privacy protection with early warning capability.

However, most existing studies commonly face two prominent problems: the lack of research on agitation warnings using long-term time-series data, and the problem of imbalanced clinical data distribution. Worse still, undersampling [29] and SMOTE [30] techniques for crudely processing clinical data to mitigate imbalance problem have long been questioned by researchers. To tackle the above issues, we designed a two-stage agitation warning model based on symmetric autoencoders and a spatio-temporal learning network. The model aims to use patient data from the previous 8×24 h for next-day agitation warning research. First, multiple symmetric autoencoders are used to augment the patients' part of normal samples, and the machine learning model is pre-trained on the augmented data. Then, the remaining balanced samples are used to fine-tune the machine learning model, to prevent the performance of the deep learning models from collapsing when they are used with imbalanced datasets.

3. Methods

The architecture of the proposed SAS-Net is shown in Figure 1. SAS-Net consists of a pseudo-label data generator (Figure 1a) and a spatio-temporal learning network (STLN) (Figure 1b).

In the first stage, 80% of the normal samples are randomly selected and passed through multiple symmetric autoencoders for feature transformation and pseudo-label construction. Then, the STLN, which is composed of a convolutional network and gated multilayer perception, is pre-trained to learn the intrinsic structure, internal rules, and effective feature extraction methods of the patient data. It is worth noting that the STLN of SAS-Net is a four-class classification network in this first stage. In the second stage, the parameters of the STLN's backbone network are frozen. We modify the fully connected classifier of STLN (from four classes to two classes) and fine-tune its parameters based on a balanced dataset (the remaining 20% of normal samples and all agitated samples). Finally, a two-class STLN is obtained for predicting whether agitation will occur or not.

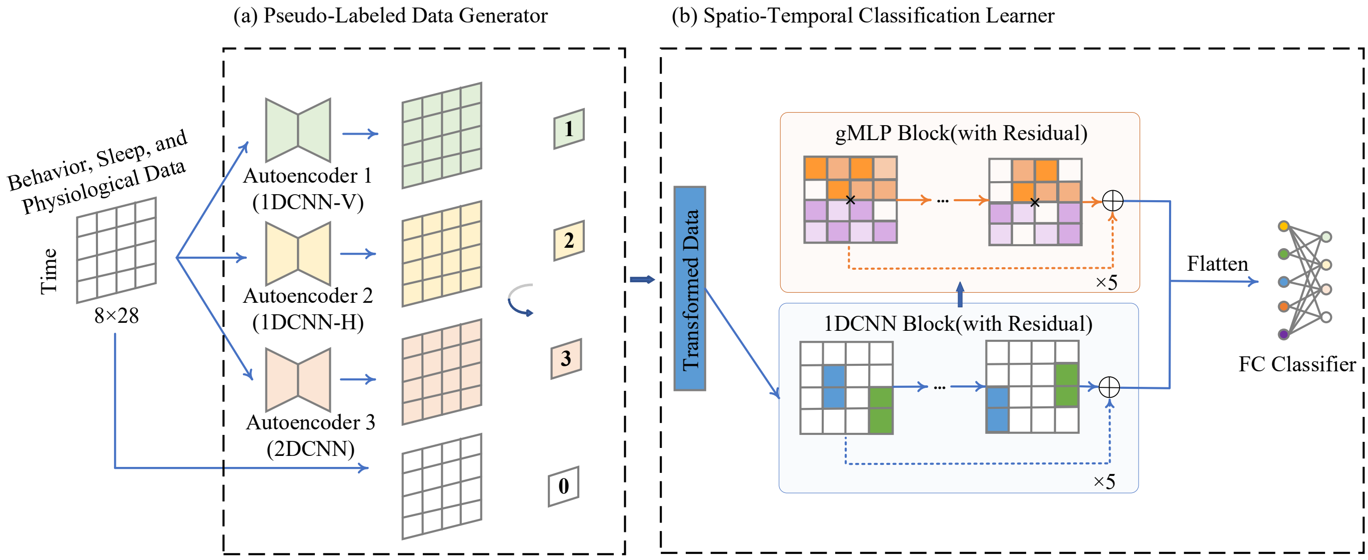


Figure 1. Network architecture of SAS-Net.

3.1. The First Training Stage

The first training stage consists of two modules: the pseudo-label data generator and the spatio-temporal learning network (STLN). The pseudo-label data generator is responsible for transforming part of the multi-modal normal sample data (including patient behavior, sleep, and physiological data). The STLN is responsible for learning the initial parameters of the backbone network.

The pseudo-label data generator is composed of several symmetric-structured autoencoders. Each autoencoder consists of an encoder and a decoder, where the decoder is built by inverting the parameters of the encoder, for ensuring that the input and output data's shapes are consistent. The pseudo-label data generator contains three types of autoencoder structures (Figure 2 shows one case of these symmetric autoencoder structures). The first one uses one-dimensional convolution to extract vertical temporal features from the data (dubbed 1DCNN-V), aiming to improve the ability of the warning network to perceive temporal changes. 1DCNN-V has 6 layers, a kernel size of 3, and a middle vector dimension of 32, and the pseudo-label of the transformed data is set to 1. The second one uses one-dimensional convolution to extract horizontal attribute features from the signals (dubbed 1DCNN-H), aiming to enhance the ability of the warning network to capture relationships between attributes. 1DCNN-H has the same number of layers, kernel size, and middle vector dimension as 1DCNN-V, and the pseudo-label of the transformed data is set to 2. The third one uses two-dimensional convolution to extract features from the data (dubbed 2DCNN), giving the warning network the ability to learn global features. 2DCNN has 6 layers, a kernel size of 3×3, a flattened middle vector dimension of 32, and the pseudo-label is set to 3. The pseudo-label data generator also retains the original untransformed data and labels it as 0; thus, the pre-training stage involves a four-class classification task.

Based on transformed data, SAS-Net uses STLN to jointly extract multi-modal behavioral, sleep, and physiological features.

As shown in Figure 1b, STLN first uses five serial 1DCNN [13] blocks to extract temporal information from patient data. Each 1DCNN block consists of a convolutional layer, a pooling layer, a normalization layer, and a dropout layer. The normalization layer is used to eliminate scale differences between features. The dropout layer is used to reduce model overfitting. At the same time, to reduce network redundancy, we set up residual connections before and after the 1DCNN modules.

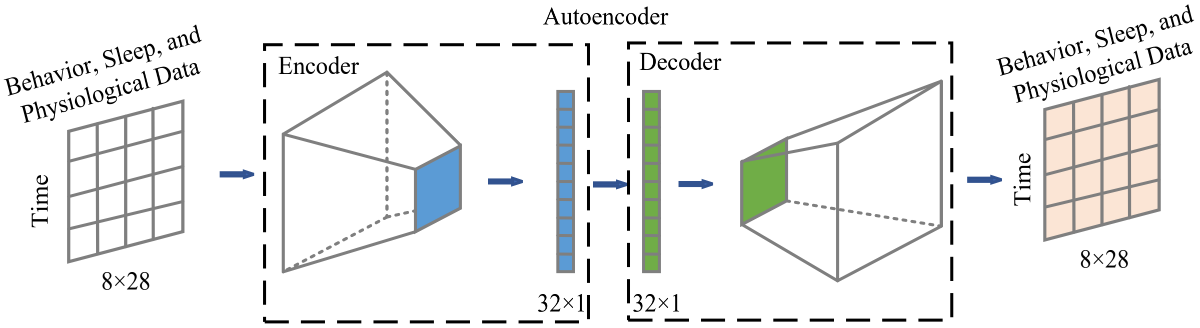


Figure 2. A symmetric autoencoder.

The gated Multilayer Perceptron (gMLP) [14] has been shown to have a strong spatial feature extraction ability. Based on the features extracted using 1DCNN, we further use five gMLP blocks to jointly extract spatial information. Here, the 1DCNN features are denoted as $X \in R^{n \times \mathrm{d}}$, where n represents the number of attribute channels and d represents the feature dimension. Each gMLP block is defined as follows:

$$G_1 = \sigma(W_1 X + b_1) \tag{1}$$

$$G_2 = sgu(G_1) \tag{2}$$

$$G_3 = W_2 X + b_2 \tag{3}$$

where G_1 and G_3 represent MLP layers, which are used to highlight the important features of different attribute data (sleep, physiological, and behavioral data). σ denotes the GELU activation function, W_1 and W_2 represent channel mappings, and b_1 and b_2 are two sets of bias values. G_2 is a spatial gating unit, which makes an original feature vector with spatial information and is the core of the entire gMLP block. Its calculation formula is as follows:

$$sgu = G_{11} \odot (G_{12} W_3 + b_3) \tag{4}$$

where G_{11} and G_{12} are two groups of features obtained by splitting the mapped feature G_{12} along the channel dimension. One of the groups, G_{12}, learns the strength of position interactions among different categories of features through spatial mapping. Then, the position interaction strength is used as a weight and applied to the first group of features through element-wise multiplication ($\odot$), thereby giving the features spatial information. Similarly, to reduce network feature redundancy, we also set up residual connections before and after the gMLP modules.

After performing a flattening operation on the spatio-temporal joint features above, the model uses a fully connected layer classifier to achieve classification on samples with different pseudo-labels. The specific parameter settings of the spatio-temporal learning network are listed in Table 1.

Table 1. Architecture and parameter settings of spatio-temporal learning network (STLN).

Blocks	Layers	Kernel Size/Stride/Padding	Input Size
1DCNN1	Conv1 Pool1	$1 \times 5/1 \times 1/0 \times 2$ $1 \times 2/1 \times 2$	$8 \times 28 \times 1$ $8 \times 28 \times 8$
1DCNN2-5 (repeat 4 times)	Conv2(2) Pool2(5)	$1 \times 3/1 \times 1/0 \times 1$ $1 \times 2/1 \times 2$	$8 \times 14 \times 16$ $8 \times 1 \times 28$
-	view	remove dimension two	$14 \times 1 \times 28$

Table 1. *Cont.*

Blocks	Layers	Kernel Size/Stride/Padding	Input Size
gMLP1-5 (repeat 5 times)	-	hidden:512	14 × 28
-	Mean	-	14 × 28
-	FC	neurons:512	1 × 28
-	-	-	1 × 4

3.2. The Second Training Phase

During the second training stage, we freeze the backbone network of the pre-trained STLN (the network structure in the dashed box in Figure 3). During the pre-training process on large-scale multi-modal data, this backbone network has already fully learned general feature extraction methods for multi-modal samples. Its parameters and weights already possess the ability to extract features from data of dementia patients, and therefore do not need to be updated again in the downstream classification task.

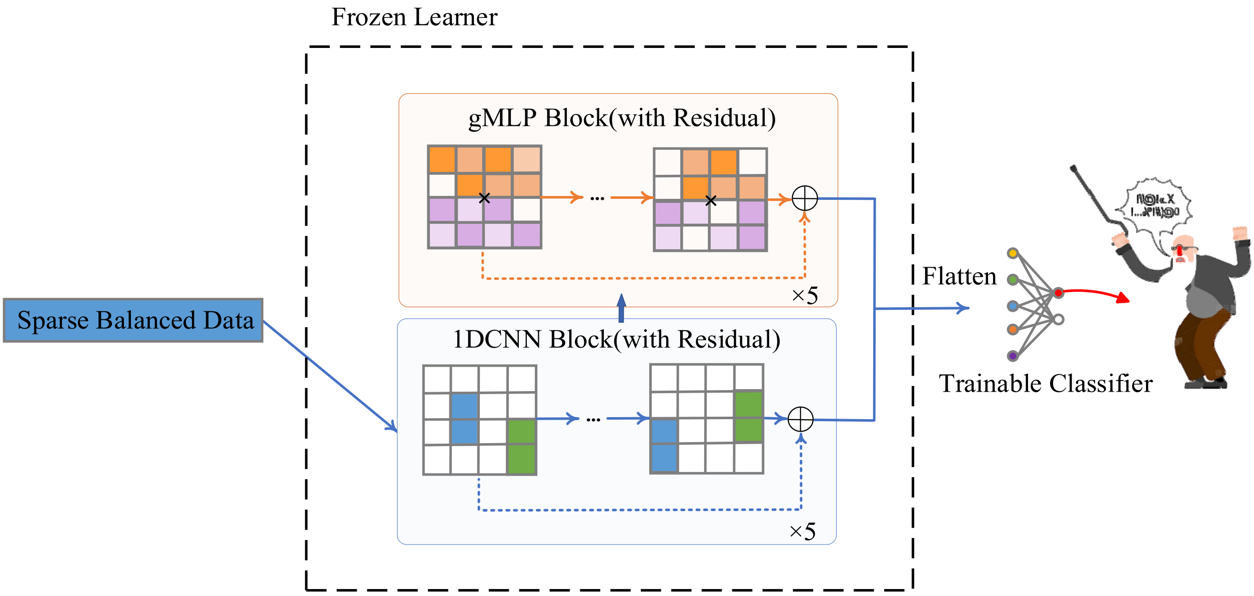


Figure 3. The second training stage.

On this basis, we use the remaining 20% of normal samples together with the sparse agitation samples, only make lightweight modifications to the trainable classifier (changing from 4 classes to 2 classes), and fine-tune the classifier parameters in the network, in order to obtain an agitation warning network under the class-imbalanced dataset.

3.3. Loss Function

In this paper, we choose the cross-entropy loss function as the optimization objective function. The cross-entropy loss function can effectively measure the difference between two probability distributions and is particularly suitable for classification problems. Its calculation formula is as follows:

$$\ell_{cross-entropy} = -[y log(\hat{y}) + (1-y) log(1-\hat{y})] \tag{5}$$

Here, y is the true distribution, and $\hat{y}$ is the class probability predicted by the model. Compared with the mean squared error loss function, cross-entropy has the advantages of stable gradients and faster convergence speed in classification tasks, and can alleviate the problem of vanishing gradients.

4. Experiment

4.1. Dataset

This study conducts experiments using the TIHM [15] dataset, an open access dataset for remote health monitoring of dementia. This dataset is a multi-modal time-series dataset for home health monitoring of dementia patients, containing data on home activities, sleep, physiological indicators, and health event labels. The dataset includes 56 patients (female: male = 28:28; age groups (70–80): (80–90): (90–100) = 17:26:13) diagnosed with dementia (Alzheimer's disease: Vascular Dementia: other dementia: N/A = 40:6:9:1), with an average of 50 days of valid data per patient, totaling 2803 days of monitoring. The "activity" table records activity perception data from multiple door magnetic sensors placed around their home, where smart home sensor layouts may differ across individuals, yet the dataset maintains uniformity in both the categories and number of sensors for all subjects. The "sleep" table records the sleep stage, heart rate, respiration, and other information on the subjects collected by the sleep pad. The "physiology" table includes daily physiological indicators for the patients, such as blood pressure, body temperature, and weight etc. The "labels" table provides clinical validated health abnormality labels for six categories: agitation, blood pressure, body temperature, weight, pulse, and body water (in this study, we only used the agitation label). The detailed information for each data table is shown in Table 2.

Table 2. Statistical information of the TIHM dataset.

Data	Number	Time Granularity
activity	1,030,560	min
sleep	2,768,544	min
physiology	17,680	once a day/several times a day
labels	1827	once a day/several times a day

Clearly, the above data have issues with inconsistency and misalignment for both collection time points and time granularities. To address this, we aggregated behavioral activity data, sleep data, and physiological data on a daily basis (by natural day). The aggregation data included the daily trigger count of each behavioral sensor, the mean and variance of heart rate and respiratory rate during sleep, and the daily mean and daily variance of multiple physiological indicators such as weight, blood pressure, and muscle mass, etc. For the 16% missing values, we adopt adjacent data for imputation. Finally, we extracted continuous eight-day, 28-dimensional time-series data as model training samples, and matched them with corresponding agitation labels for subsequent research and construction of the agitation warning model. Detailed information on the processed data is shown in Table 3.

Table 3. Detailed information on the processed data.

Number of Features	Number of Labels	Number of Agitation Samples	Number of Normal Samples
2149	2149	303	1846

4.2. Experimental Setup

The parameters of the proposed SAS-Net are as follows: the learning rate is set to 1×10^{-3}, the batch size is set to 64, and the number of training epochs is set to 200.

For the optimization strategy, the Adam optimizer is used to iteratively optimize the network parameters.

The experiments were conducted on an NVIDIA RTX 4060 GPU with 8 GB of memory, using CUDA 12.7 for GPU acceleration, and the model was built and implemented in the Python 3.10 development environment.

In the experimental phase (Tables 4 and 5), remaining balanced subject data were pooled. Agitated and normal samples were split at a ratio of 2:8 independently, with 20% of each category assigned to the test set and the remaining 80% to the training set. Based on this partitioning strategy, 5-fold cross-validation was utilized to evaluate the models.

4.3. Evaluation Metrics

We select Accuracy, Precision, Recall, and F1 Score as the evaluation metrics. Let the set of positive samples (agitation samples) be P, and the set of true positive samples be T. TP is the number of samples that are both predicted as positive and are actually positive. FP is the number of samples predicted as positive but are actually negative. TN is the number of samples predicted as negative and are actually negative. FN is the number of samples predicted as negative but are actually positive.

(1) Accuracy

Accuracy is a basic indicator for measuring the overall correctness of the model's warnings. It reflects the proportion of correctly classified samples among all samples and intuitively shows the overall classification ability of the model. Its calculation formula is

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

(2) Precision

Precision is used to measure the proportion of true positive samples among the samples predicted as positive by the model. It reflects the Accuracy of the model's positive warnings and can effectively evaluate the model's ability to avoid mistakenly identifying a patient as agitated. Its calculation formula is

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

(3) Recall

Recall is used to measure the proportion of true positive patients who are successfully predicted as positive by the model. It reflects the model's ability to cover positive samples. In medical scenarios, it can effectively avoid missing the risk of agitation in patients, and is one of the core indicators of concern in this study. Its calculation formula is

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

(4) F1 Score

The F1 Score is the harmonic mean of Precision and Recall. It is used to balance the model's Precision and coverage capabilities, avoiding the one-sidedness of a single indicator. It can comprehensively reflect the model's classification performance on imbalanced data and is particularly suitable for the medical scenario in this study. Its calculation formula is

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

4.4. Comparative Experiments

To verify the effectiveness of our model, we compared SAS-Net with several mainstream deep learning networks in a first comparative experiment. The specific configuration details of each comparison model are as follows:

(1) SA-LSTM and SA-GRU: These two networks follow the same basic pipeline as SAS-Net, but replace the STLN module with LSTM [31] or GRU [32] to extract temporal features. Finally, the extracted features are concatenated and passed through a fully connected layer classifier to complete the classification task.
(2) SA-2DCNN: This model replaces the STLN module with a 2DCNN [33] machine learning model network. It treats the data as a two-dimensional matrix and performs convolutional feature extraction. Then, the features are flattened and passed through a fully connected classifier to complete the classification task.
(3) SA-Transformer: This model replaces the STLN module with a Transformer [34] machine learning model. After applying positional encoding to each attribute column of data, it uses a self-attention mechanism for feature learning. Finally, the features are flattened and passed into a fully connected layer for classification.
(4) SMOTE-SNet: This model directly uses the SMOTE technique to augment samples, and then employs the spatio-temporal network proposed in this paper for agitation warning. Finally, the optimal model is selected and further evaluated on the sparse sample set.

Table 4 shows the performance comparison of different deep learning networks. The results show that the SA-LSTM and SA-GRU models suffer from serious performance issues. SA-LSTM achieves zero for Precision, Recall, and F1 Score, and SA-GRU also has a low Recall and F1 Score, indicating that these two types of temporal recurrent networks fail to capture effective features from the data and therefore have difficulty completing the agitation warning task. Although the SA-2DCNN and SA-Transformer models achieve basic performances, with F1 Score improved to 0.6667, their Recall performance is insufficient and poses a high risk of missed detection. Meanwhile, SMOTE-SNet achieves a lower Precision due its the crude data augmentation technique. SAS-Net, our model, achieves balanced and leading advantages across multiple indicators: Accuracy reaches 0.9391, which is 4.92% higher than SA-Transformer. The F1 Score reaches 0.8293, a significant improvement of 24.48% over the baseline model.

Table 4. Performance comparison under different deep learning networks.

Model	Accuracy	Precision	Recall	F1 Score
SA-LSTM	0.8087	0	0	0
SA-GRU	0.8261	1	0.0909	0.1667
SA-2DCNN	0.8957	0.8571	0.5455	0.6667
SA-Transformer	0.9043	1	0.5000	0.6667
SMOTE-SNet	0.8000	0.4889	1	0.6567
SAS-Net	0.9391	0.8947	0.7727	0.8293

Figure 4 shows the confusion matrices of warning networks of different deep learning networks. The confusion matrices provide an intuitive comparison of the classification performance of each model in the agitation warning task. Among them, SA-LSTM and SA-GRU show severe classification bias. SA-LSTM misclassifies all agitation samples as normal samples and completely fails to identify agitation events. SA-GRU correctly classifies only two agitation samples, with a very high rate of missed detections. SA-2DCNN and SA-Transformer can identify some agitation samples, but still have many missed detections, resulting in limited classification performances. SMOTE-SNet suffers from overfitting, which triggers severe agitation events false alarms (there were 23 normal samples misidentified as agitated cases). Our proposed SAS-Net achieves a well-balanced performance, attaining a high classification Accuracy and significantly outperforming the comparative models. The experimental results above fully validate the classification robustness and performance of SAS-Net on imbalanced data.

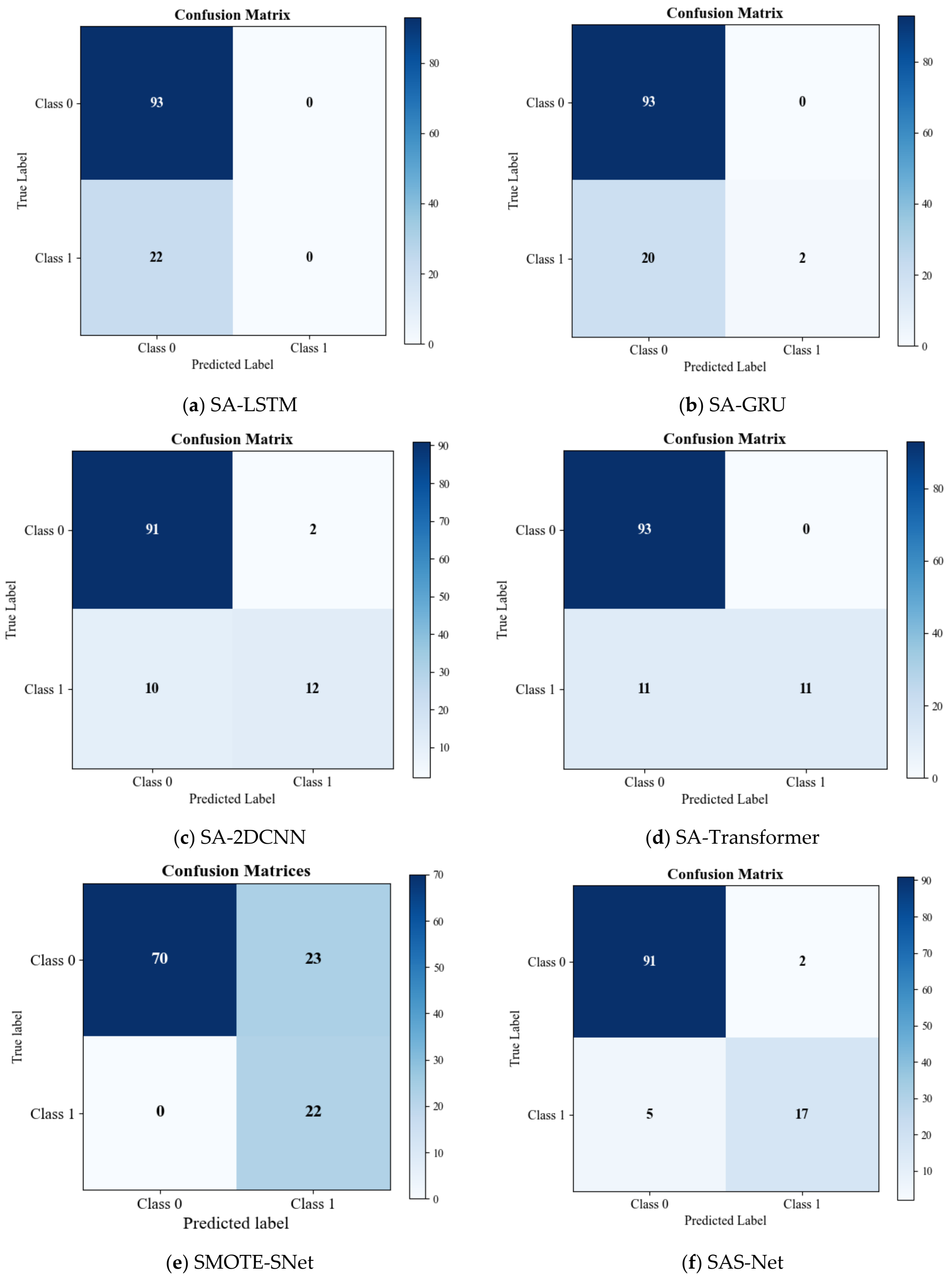


Figure 4. Confusion matrices of different deep learning networks.

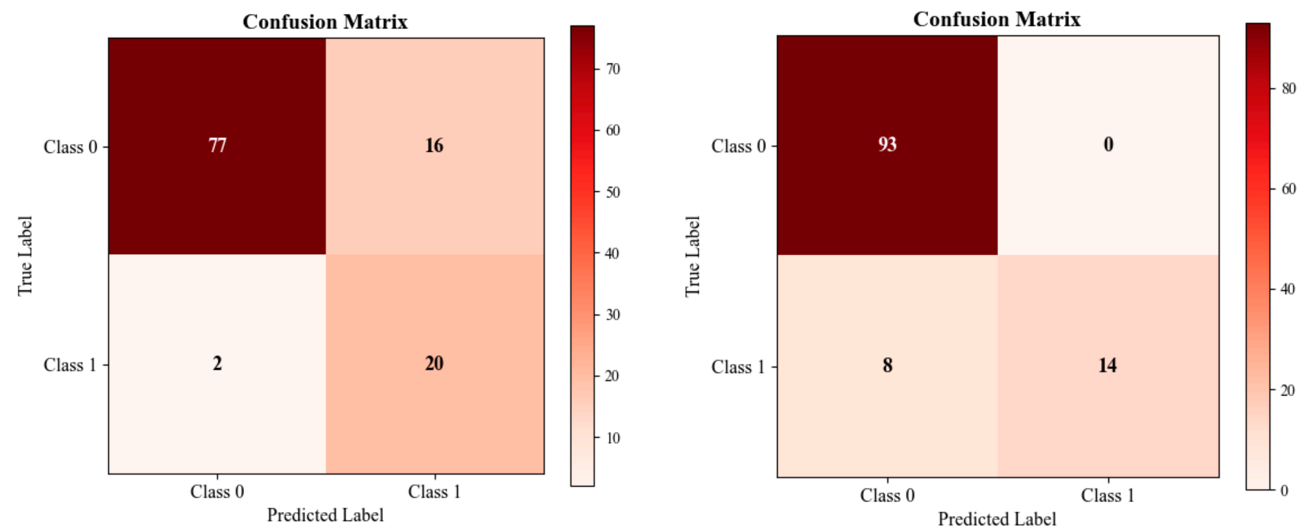
To further verify the performance of SAS-Net, we conducted another comparative experiment. The symmetric autoencoders and the machine learning model network of SAS-Net were kept unchanged, and only the classifier of SAS-Net was replaced. In Table 5, it can be seen that SAS-Bayes, SAS-KNN, SAS-Decision Tree, SAS-XGBoost, and SAS-LightGBM replace the fully connected layer classifiers of SAS-Net with Naive Bayes, K-Nearest Neighbors, Random Forest, XGBoost, and LightGBM, respectively.

Table 5. Comparative experiments using different classifiers.

Model	Accuracy	Precision	Recall	F1 Score
SAS-Bayes	0.8435	0.5556	0.9091	0.6897
SAS-KNN	0.9304	1	0.6364	0.7778
SAS-Decision Tree	0.9304	0.8182	0.8182	0.8182
SAS-XGBoost	0.9304	0.8500	0.7727	0.8095
SAS-LightGBM	0.9391	0.9412	0.7273	0.8205
SAS-Net	0.9391	0.8947	0.7727	0.8293

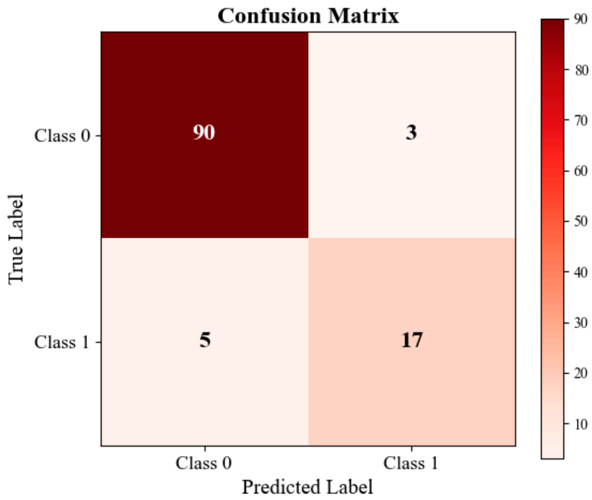
Table 5 presents the results of using different classifiers. The experiments show that SAS-Bayes achieves the highest Recall rate, but its Precision and F1 Score are relatively low, resulting in a large number of false positives. SAS-KNN achieves a Precision of 1.0, but its low Recall rate leads to a high risk of missed detections. SAS-Decision Tree, SAS-XGBoost, and SAS-LightGBM demonstrate balanced performances, but most of their metrics are inferior to those of SAS-Net.

Figure 5 shows the confusion matrices under different classifiers. SAS-Bayes has a very high false positive rate. SAS-KNN has eight missed detections. Although SAS-Decision Tree, SAS-XGBoost, and SAS-LightGBM perform steadily, they are slightly inferior to SAS-Net. From Figures 4f and 5, it can be seen once again that SAS-Net performs the best, achieving a good balance between false positives and missed detections.

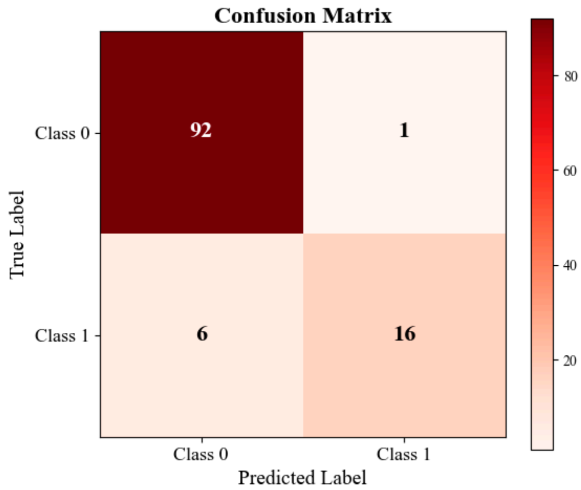


(**a**) SAS-Bayes (**b**) SAS-KNN

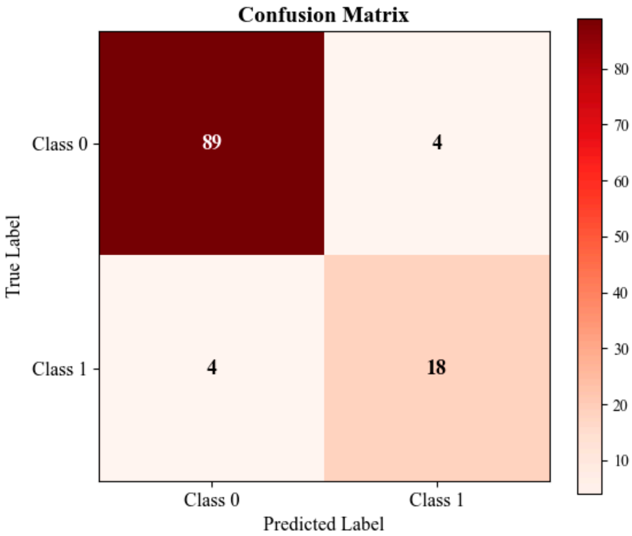
Figure 5. *Cont.*



(**c**) SAS-XGBoost



(**d**) SAS-LightGBM



(**e**) SAS-Decision Tree

Figure 5. Confusion matrices under different classifiers.

4.5. Ablation Experiments

To verify the effectiveness of each component of SAS-Net, we conducted ablation experiments on core components of SAS-Net. First, we removed the pseudo-label data generator based on symmetric autoencoders to explore the contribution of the pre-trained process to model performance. This model is named (w/o)SA. This model does not require pre-training. Second, we ablated the 1DCNN block of STLN to verify the contribution of time-series feature extraction in the overall model. This model is named (w/o)T-Net. Finally, we removed the gMLP block of STLN to explore the contribution of features integrating patient behavioral, sleep, and physiological data to classification performance. This model is named (w/o)S-Net.

The ablation experiment results in Figure 6 and Table 6 fully validate the effectiveness of each core component of the SAS-Net model. After removing the pseudo-label data generator, the model's Precision, Recall, and F1 Score all dropped to zero, indicating that data augmentation is the core prerequisite for addressing data imbalance and achieving effective identification of agitation samples. After ablating the 1DCNN block of STLN, Recall and F1 Score decreased significantly, showing that temporal feature extraction is crucial for capturing the temporal patterns of agitation events. After removing the gMLP block of STLN, the ability to fuse multi-source data was lost, and all indicators also declined markedly, validating the effectiveness of the spatial network in jointly modeling behavioral, sleep, and physiological data. The SAS-Net model achieves a comprehensive leading performance across all indicators, fully demonstrating that the three modules work together to ensure the model's excellent performance in the agitation warning task.

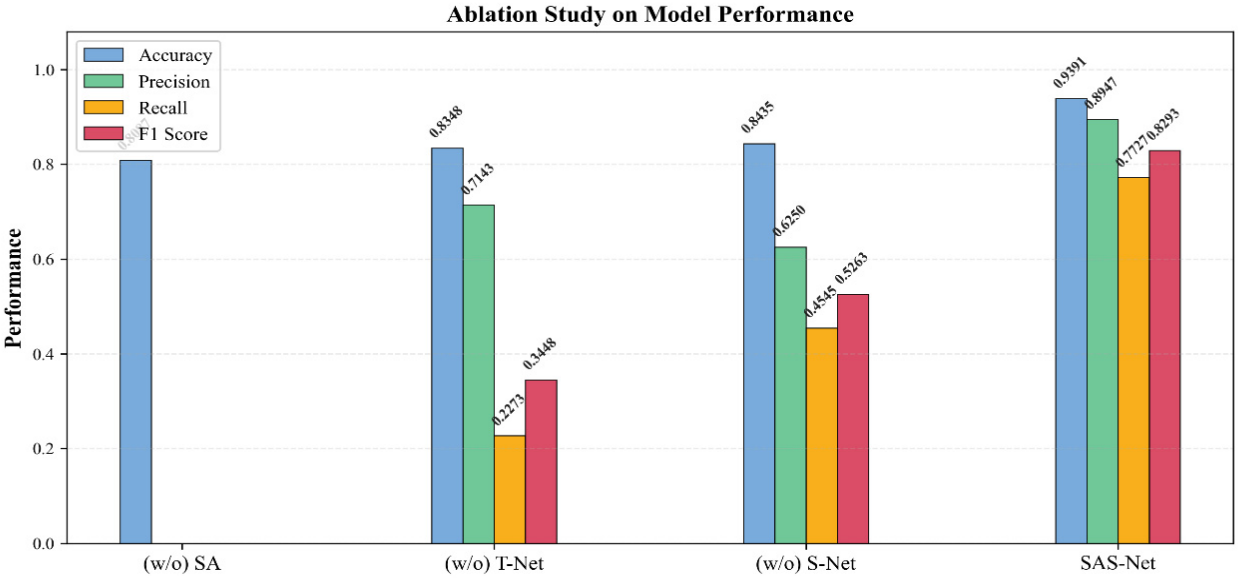


Figure 6. Ablation experiment statistics.

Table 6. Ablation experiments.

Model	Accuracy	Precision	Recall	F1 Score
(w/o) SA	0.8087	0	0	0
(w/o) T-Net	0.8348	0.7143	0.2273	0.3448
(w/o) S-Net	0.8435	0.6250	0.4545	0.5263
SAS-Net	0.9391	0.8947	0.7727	0.8293

5. The Number of Symmetric Autoencoders

This section discusses the impact of the number of symmetric autoencoders on model performance. The experimental design is shown in Table 7. We set the number of symmetric autoencoders to 1, 2, 3, 4, 5, and 6. In Table 7, ∣SA∣ denotes the number of symmetric autoencoders, and the subscript 16 indicates a variant structure where the autoencoder's middle vector dimension is set to 16. Then, under the same experimental configuration, we trained the SAS-Net model and evaluated the impact of the number of symmetric autoencoders on the performance from four indicators (Accuracy, Precision, Recall, and F1 Score).

Table 7. Experimental design of the number of symmetric autoencoders.

∣SA∣	Symmetric Autoencoders
1	1DCNN-V
2	1DCNN-V, 1DCNN-H
3	1DCNN-V, 1DCNN-H, and 2DCNN
4	1DCNN-V, 1DCNN-H, 2DCNN, 1DCNN-V_{16}
5	1DCNN-V, 1DCNN-H, 2DCNN, 1DCNN-V_{16}, and 1DCNN-H_{16}
6	1DCNN-V, 1DCNN-H, 2DCNN, 1DCNN-V_{16}, 1DCNN-H_{16}, and $2DCNN_{16}$

The experimental results are shown in Table 8 and Figure 7. When |SA| = 1, the model achieves a Precision of 1.0, but Recall is insufficient, posing a high risk of missed detection. When |SA| = 2, the model performance becomes balanced. When |SA| = 3, the model performs excellently, with Accuracy reaching 0.9391 and F1 Score reaching 0.8293, both of which are the highest in this series of experiments, indicating that the warning network achieves a good balance between false positives and missed detections at this setting. When |SA| > 3 (i.e., the number of autoencoders is 4, 5, or 6), model performance shows an overall downward trend. At |SA| = 4, although Recall slightly increases to 0.8182, Accuracy, Precision, and F1 Score drop to 0.9130, 0.7500, and 0.7826, respectively. At |SA| = 5 and |SA| = 6, all indicators drop significantly, and we speculate that the core reason for this performance decline is that too many augmented categories make the pre-training multi-class classification task much more difficult, making it hard to effectively guide the downstream binary recognition task, thereby reducing generalization ability and Accuracy. Finally, considering both model performance and computational efficiency, we selected three symmetric autoencoders for the pseudo-label data generator.

Table 8. Influence of number of symmetric autoencoders on model performance.

Model	Accuracy	Precision	Recall	F1 Score
SAS-Net$_{(|SA|=1)}$	0.9217	1	0.5909	0.7429
SAS-Net$_{(|SA|=2)}$	0.9130	0.8000	0.7273	0.7619
SAS-Net$_{(|SA|=3)}$	0.9391	0.8947	0.7727	0.8293
SAS-Net$_{(|SA|=4)}$	0.9130	0.7500	0.8182	0.7826
SAS-Net$_{(|SA|=5)}$	0.8870	0.7368	0.6364	0.6829
SAS-Net$_{(|SA|=6)}$	0.8261	0.5417	0.5909	0.5652

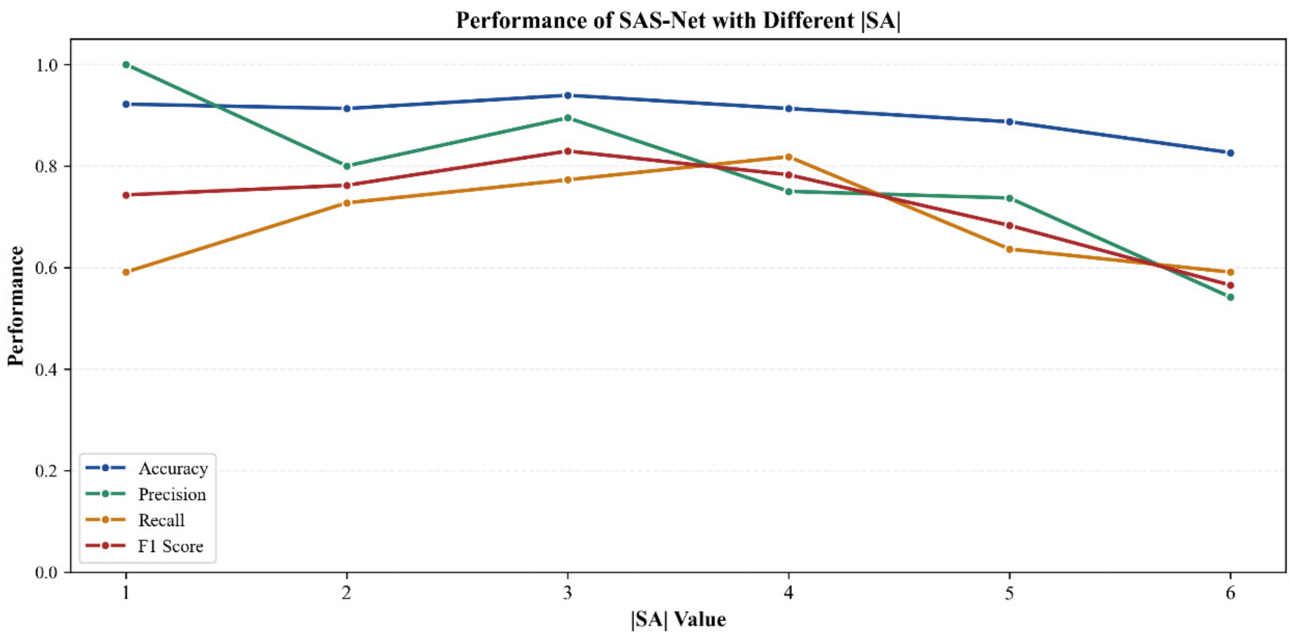


Figure 7. Model performance under different numbers of symmetric autoencoders.

6. Conclusions

To address the two prominent problems, the insufficient research on agitation warning using long-term time-series data and the extreme imbalance problem in clinical data, we built an agitation warning network for dementia patients using 8 × 24 h data based on symmetric autoencoders and a spatio-temporal learning network, dubbed SAS-Net. To solve the data imbalance problem, this study designs a two-stage training strategy of "machine learning model construction–machine learning model fine-tuning." First, multiple symmetric autoencoders are employed to augment partial normal samples of patients, and the machine learning model is pre-trained on the augmented data. Then, the remaining balanced samples are used to fine-tune the machine learning model, which effectively prevents the deep learning model's performance from collapsing when it is used with imbalanced datasets. In the experimental stage, two sets of comparative experiments were conducted. In the first set, we compared our model with various state-of-the-art machine learning networks. In the second set, the fully connected layer classifier was replaced with general classifiers for competition. These results show that the proposed model achieves a balanced and leading performance in terms of many indicators. Finally, ablation experiments removed the symmetric autoencoders, the temporal network, and the spatial network to verify the effectiveness of each core component, and the SAS-Net achieves the best performance on all metrics, fully demonstrating the rationality of the proposed model. While the two-stage SAS-Net framework has obtained modest achievements, the extreme sparse agitation samples remain a foundational learning challenge. In future work, we will continue to address the problem by introducing data augmentation and few-shot learning algorithms, etc.

Author Contributions: Conceptualization, methodology, J.X.; software, validation, visualization, B.L.; writing—original draft preparation, writing—review and editing, P.F., Y.Z. and S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by Key Laboratory of Intelligent Rehabilitation and Barrier-free for the Disabled (Changchun University), Ministry of Education and the Research project of Jilin Provincial Department of Education under grant no JJKH20251100KJ.

Data Availability Statement: Data is available in a publicly accessible repository. The TIHM dataset presented in this study is openly available at: https://github.com/PBarnaghi/TIHM-Dataset (accessed on 10 June 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Scheltens, P.; De Strooper, B.; Kivipelto, M.; Holstege, H.; Chételat, G.; Teunissen, C.E.; Cummings, J.; Van der Flier, W.M. Alzheimer's disease. *Lancet* **2021**, *397*, 1577–1590. [PubMed]
2. Black, V.; Arteaga, I.; Friend, J. Leveraging the Dyad to Inform Dementia Care Policy: Some Anthropological Precautions and Possibilities. *Am. J. Bioeth.* **2026**, *26*, 77–79. [CrossRef] [PubMed]
3. Chiao, C.Y.; Wu, H.S.; Hsiao, C.Y. Caregiver burden for informal caregivers of patients with dementia: A systematic review. *Int. Nurs. Rev.* **2015**, *62*, 340–350. [CrossRef] [PubMed]
4. Bruvik, F.K.; Ulstein, I.D.; Ranhoff, A.H.; Engedal, K. The quality of life of people with dementia and their family carers. *Dement. Geriatr. Cogn. Disord.* **2012**, *34*, 7–14. [CrossRef] [PubMed]
5. Farina, N.; Page, T.E.; Daley, S.; Brown, A.; Bowling, A.; Basset, T.; Livingston, G.; Knapp, M.; Murray, J.; Banerjee, S. Factors associated with the quality of life of family carers of people with dementia: A systematic review. *Alzheimer's Dement.* **2017**, *13*, 572–581. [CrossRef]
6. Spasojevic, S.; Nogas, J.; Iaboni, A.; Ye, B.; Mihailidis, A.; Wang, A.; Li, S.J.; Martin, L.S.; Newman, K.; Khan, S.S. A pilot study to detect agitation in people living with dementia using multi-modal sensors. *J. Healthc. Inform. Res.* **2021**, *5*, 342–358. [CrossRef] [PubMed]
7. Bafaloukou, M.; Schalkamp, A.K.; Fletcher-Lloyd, N.; Capstick, A.; Walsh, C.; Sandor, C.; Kouchaki, S.; Nilforooshan, R.; Barnaghi, P. An interpretable machine learning tool for in-home monitoring of agitation episodes in people living with dementia: A proof-of-concept study. *eClinicalMedicine* **2025**, *80*, 103032. [PubMed]
8. Abedi, A.; Chu, C.H.; Khan, S.S. Early Prediction of Agitation in Community-Dwelling People with Dementia Using Multimodal Sensors and Machine Learning: Benchmarking of State-of-the-Art Techniques. In Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 16–22 August 2025.

9. Logsdon, R.G.; Teri, L.; Weiner, M.F.; Gibbons, L.E.; Raskind, M.; Peskind, E.; Grundman, M.; Koss, E.; Thomas, R.G.; Thai, L.J. Assessment of agitation in Alzheimer's disease: The agitated behavior in dementia scale. *J. Am. Geriatr. Soc.* **1999**, *47*, 1354–1358. [CrossRef] [PubMed]
10. Van der Mussele, S.; Le Bastard, N.; Saerens, J.; Somers, N.; Mariën, P.; Goeman, J.; De Deyn, P.P.; Engelborghs, S. Agitation-associated behavioral symptoms in mild cognitive impairment and Alzheimer's dementia. *Aging Ment. Health* **2015**, *19*, 247–257. [PubMed]
11. Mega, M.S.; Cummings, J.L.; Fiorello, T.; Gornbein, J. The spectrum of behavioral changes in Alzheimer's disease. *Neurology* **1996**, *46*, 130–135. [PubMed]
12. Li, X.; Hu, N.; Tan, M.S.; Yu, J.T.; Tan, L. Behavioral and psychological symptoms in Alzheimer's disease. *BioMed Res. Int.* **2014**, *2014*, 927804. [CrossRef] [PubMed]
13. Kim, Y. Convolutional neural networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 26–28 October 2014.
14. Liu, H.; Dai, Z.; So, D.; Le, Q.V. Pay attention to mlps. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Online, 6–14 December 2021.
15. Palermo, F.; Chen, Y.; Capstick, A.; Fletcher-Loyd, N.; Walsh, C.; Kouchaki, S.; True, J.; Balazikova, O.; Soreq, E.; Scott, G.; et al. TIHM: An open dataset for remote healthcare monitoring in dementia. *Sci. Data* **2023**, *10*, 606. [CrossRef] [PubMed]
16. Deters, J.K.; Janus, S.; Silva, J.A.L.; Wörtche, H.J.; Zuidema, S.U. Sensor-based agitation prediction in institutionalized people with dementia A systematic review. *Pervasive Mob. Comput.* **2024**, *98*, 101876. [CrossRef]
17. Moore, P.; Thomas, A.; Tadros, G.; Xhafa, F.; Barolli, L. Detection of the onset of agitation in patients with dementia: Real-time monitoring and the application of big-data solutions. *Int. J. Space-Based Situated Comput.* **2013**, *3*, 136–154.
18. Fletcher-Lloyd, N.; Soreq, E.; Wilson, D.; Nilforooshan, R.; Sharp, D.J.; Barnaghi, P. Home monitoring of daily living activities and prediction of agitation risk in a cohort of people living with dementia. *Alzheimer's Dement.* **2021**, *17*, e058614. [CrossRef]
19. Umah, B.; Palaniswamy, S.; Kumaran, U. Predicting Agitation in Dementia Patients Using IoT Sensor Data and Deep Learning Techniques. In Proceedings of the International Conference on Cloud Computing, Data Science and Engineering, Uttar Pradesh, India, 16–17 January 2025.
20. Hasan, Z.; Shum, L.C.; Faruk, T.; McGilton, K.S.; Chu, C.H.; McArthur, C.; Flint, A.J.; Mihailidis, A.; Khan, S.S.; Iaboni, A. Machine learning to detect motor agitation in people with dementia using real-time location data. *Am. J. Geriatr. Psychiatry* **2025**, *33*, 1171–1183. [CrossRef] [PubMed]
21. Khan, S.S.; Mishra, P.K.; Ye, B.; Patel, S.; Newman, K.; Mihailidis, A.; Iaboni, A. A novel multi-modal sensor dataset and benchmark to detect agitation in people living with dementia in a residential care setting. *ACM Trans. Comput. Healthc.* **2025**, *6*, 1–19. [CrossRef]
22. Ramesh, K.; Yakoub, A.; Ghoneim, Y.; Al Korabi, R.; Ramesh, J.; Sagahyroon, A.; Aloul, F. Machine Learning Prediction of Agitation in Dementia Patients Using Sleep and Physiological Data. *Appl. Sci.* **2025**, *15*, 9908. [CrossRef]
23. Malioukis, A.; Snead, R.S.; Marczika, J.; Ambalavanan, R.; Towett, G.; Mbogori-Kairichi, M. Digital biomarkers for early agitation detection in dementia: A scoping review of emerging wearable and smart technologies for personalized care. *Front. Neurol.* **2026**, *17*, 1683517. [CrossRef] [PubMed]
24. Munteanu, D.; Bejan, C.; Munteanu, N.; Zamfir, C.; Vasić, M.; Petrea, S.M.; Cristea, D. Deep-learning-based system for assisting people with Alzheimer's disease. *Electronics* **2022**, *11*, 3229.
25. Rezvani, R.; Kouchaki, S.; Nilforooshan, R.; Sharp, D.J.; Barnaghi, P. Semi-supervised learning for identifying the likelihood of agitation in people with dementia. *arXiv* **2021**, arXiv:2105.10398.
26. Pragadeeswaran, S.; Mohankumar, N.; Puviarasi, R.; Nandhini, N.; Sivaraman, A.; Rajmohan, M. Detecting and Predicting Agitation Episodes in Dementia Patients Using IoT and Artificial Neural Networks. In Proceedings of the 2024 International Conference on Emerging Technologies and Innovation for Sustainability (EmergIN), Greater Noida, India, 20–21 December 2024.
27. HekmatiAthar, S.; Goins, H.; Samuel, R.; Byfield, G.; Anwar, M. Data-driven forecasting of agitation for persons with dementia: A deep learning-based approach. *SN Comput. Sci.* **2021**, *2*, 326. [CrossRef] [PubMed]
28. Badawi, A.; Elmoghazy, S.; Choudhury, S.; Elgazzar, S.; Elgazzar, K.; Burhan, A.M. Multimodal Detection of Agitation in People with Dementia in Clinical Settings: Observational Pilot Study. *JMIR Aging* **2025**, *8*, e68156. [CrossRef] [PubMed]
29. Meng, Z.; Iaboni, A.; Ye, B.; Newman, K.; Mihailidis, A.; Deng, Z.; Khan, S.S. Undersampling and cumulative class re-decision methods to improve detection of agitation in people with dementia. *Biomed. Eng. Lett.* **2024**, *14*, 69–78. [PubMed]
30. Charte, F.; Rivera, A.J.; Del Jesus, M.J.; Herrera, F. MLSMOTE: Approaching imbalanced multilabel learning through synthetic instance generation. *Knowl.-Based Syst.* **2015**, *89*, 385–397. [CrossRef]
31. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
32. Cho, K.; Van Merriënboer, B.; Gulçehre, Ç.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 26–28 October 2014.

33. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
34. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–9 December 2017.