Multi-head CNN and LSTM with Attention for User Status Estimation from Biometric Information

Hyunseo Park
School of Electrical Engineering
Korea Advanced Institute of Science and Technology
Daejeon, Rep. of Korea, 34141
tkf92001@kaist.ac.kr

Nakyoung Kim
Institute for Information Technology Convergence
Korea Advanced Institute of Science and Technology
Daejeon, Rep. of Korea, 34141
nkim71@kaist.ac.kr

Gyeong Ho Lee
School of Electrical Engineering
Korea Advanced Institute of Science and Technology
Daejeon, Rep. of Korea, 34141
gyeongho@kaist.ac.kr

Jaeseob Han
School of Electrical Engineering
Korea Advanced Institute of Science and Technology
Daejeon, Rep. of Korea, 34141
j89449@kaist.ac.kr

Hyeontaek Oh
Institute for Information Technology Convergence
Korea Advanced Institute of Science and Technology
Daejeon, Rep. of Korea, 34141
hyeontaek@kaist.ac.kr

Jun Kyun Choi
School of Electrical Engineering
Korea Advanced Institute of Science and Technology
Daejeon, Rep. of Korea, 34141
jkchoi59@kaist.ac.kr

***Abstract*—With Internet of Things technologies, healthcare services for smart homes are emerging. In the meantime, the number of households of single-living elderly who are distant from using smart devices is increasing, and contactless radar-based sensors are recently introduced to monitor the users in single households. In this paper, contactless radar-based sensors were installed in over 100 households of single-living elderly to collect their biometric data under uncontrolled environments. In addition, a deep learning-based classification model is proposed that estimates the user status in predefined classes. In particular, the classification model is designed with a multi-head convolutional neural network with long-short-term memory and an attention mechanism. The proposed model aims to extract features in diverse resolutions from the biometric data while capturing the temporal causalities and relative importance of the features. The experimental results verify that the proposed classification model improves the status classification accuracy by 2.8% to 31.7% in terms of $\mathbf{F}_1$ score for the real-world dataset.**

***Index Terms*—Radar-based status monitoring, status classification, multi-head feature extraction**

I. INTRODUCTION

With the recent advancements of Internet of Things (IoT) technologies, various smart home applications are developed. A smart home means a residence where a household can benefit from smart services with remote access, monitoring, and control capabilities [1]. Meanwhile, the number of single households is anticipated to rise worldwide, with a significant increase in the ratio of the elderly in the next two decades [2]. This growing aging population could cause a weakening of the socio-economic structure of many countries in terms of healthcare and related costs [3]. Accordingly, many efforts to develop the solution for healthcare services for the elderly in single households (e.g., monitoring of underlying diseases, detecting emergencies, etc.) are put together.

On-body and contactless sensors have been considered as the two main approaches for healthcare services to collect and monitor users' biometric data [4]. The on-body method uses wearable devices for service provision, and the sensors need to be placed on the user's body. Hence, the biometric data collected from the on-body sensors are relatively accurate. On the other hand, the user requires to continuously wear the sensor device and frequently charge the device, which raise inconvenience especially for the high age group who are often distant from using smart devices [5]. Accordingly, when it comes to the status monitoring in the users' living space, the methods to monitor the users' status from the data collected with contactless sensors are widely investigated for service provision without affecting the users' daily life [6].

As the sensor technologies are getting mature than ever, contactless radar-based data acquisition methods are recently investigated that use the phase change of the radar reflected by the movement of the human body. As a consequence, an individual's biometric features (e.g., heart rate, respiration, etc.) can now be accurately obtained without contact, and the development of analysis on the collected biometric data is in progress [7]–[10]. In this paper, a method to estimate a user's

This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE) and Korea Institute for Advancement of Technology (KIAT) through the International Cooperative R&D program. (Project No. 0011879)

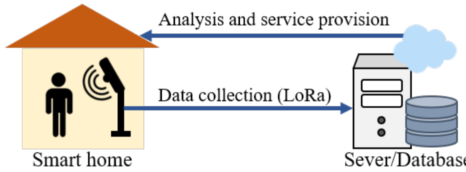

Fig. 1. High level radar-based biometric monitoring system model

Fig. 2. Illustration of the installed radar sensor

TABLE I
SPECIFICATIONS OF THE INSTALLED RADAR SENSOR

Category	Specification
Chip	Sharp DC6M4JN3000
Method	Microwaves (24.05 24.25GHz)
Resolution	60cm
Range (Max.)	1.5m (Heartbeats, Breathing) 7m (Body motion)
Directionality	Azimuth: 25°, Elevation: 20°
Error rate	±10% (3m)

status from the biometric information collected through a radar sensor is investigated. While radar sensors have a limitation that the accuracy of the measurement drops when the subject is intensely moving, this paper aims for the situations when the displacement and the motion of the subject are bound to some degree (e.g., sleeping in the bed, having a meal on the table, watching television, etc.) In the meantime, the time-series data on human activity collected from sensors can be decomposed into a set of features in multiple levels (i.e., low, middle, and high), and the means to capture the feature in diverse resolution, as well as their temporal causalities, are needed for the user status estimation from biometric data. Accordingly, a deep learning-based model is proposed in this paper that analyzes a fixed length of biometric data and classifies the user status in that period. The contactless radar-based sensors were installed in the living spaces of over 100 single households of elders, and real-world biometric information is collected. In particular, this paper utilizes the biometric information of 22 elders collected under uncontrolled environments and proposes a deep learning-based user status classification model.

The rest of this paper is organized as follows. Section II provides previous studies. Section III-B introduces the status classification method. The performance evaluation is presented in Section IV, followed by the conclusion in Section V.

TABLE II
SAMPLE DISTRIBUTION OF THE HR DATASET

Class	Status	Number of data samples	Ratio
1	Not detected	32,790	33.6%
2	In sleep	45,148	46.3%
3	In active activity	5,762	5.9%
4	In stationary Activity	6,460	6.6%
5	In unidentified activity	7,405	7.6%
Total		97,565	100%

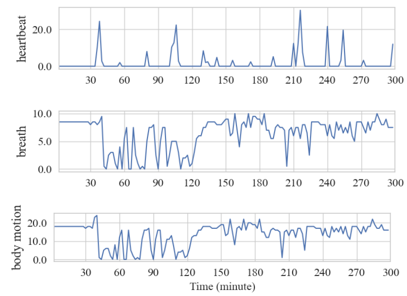

Fig. 3. Example of the biometric data of a subject in sleep

II. RELATED WORKS

The research on biometric information monitoring technologies with contactless radar-based sensors are largely in two fields of study. One is on increasing the accuracy of the contactless monitoring sensors' measurements compared to on-body sensors. The other is on improving the performance of biometric data analysis to use them for applications and services. With the recent advancements in sensor technology, the second field of study that focuses on biometric data analysis is actively in progress. In the meantime, research on radar-based cardio-respiration monitoring is widely investigated, which is mainly conducted in an impulse-radio ultra wide band (IR-UWB) radar-based environment [7]. This research includes studies that propose a method to accurately measure of the heart rate, to reliably monitor the continuous cardio-respiration rate, to estimate the sleep stages, etc. [10]–[12]. In particular, many studies have applied deep learning technologies to enhance the accuracy of the analysis on biometric data. Long-short-term memory (LSTM) with an attention mechanism is used in [12] to analyze the heart and respiratory rates collected by an IR-UWB radar-based device. In [13] and [14], electrocardiogram signals are analyzed respectively based on a multi-head and residual convolutional neural network (CNN). In the meantime, the well-known deep learning technologies are applied to a set of publicly available biometric datasets, and their performances are compared and evaluated in [15]. However, to the best knowledge of the authors, there has not been a work that comprehensively utilizes multi-head CNN with LSTM and an attention mechanism to enhance the performance of the user's status classification. Accordingly, this study proposes a deep learning-based user status classification model designed with multi-head CNNs, LSTMs, and a multi-head self-attention mechanism to integrally capture the contexts in diverse feature levels as well as their temporal causalities and importance.

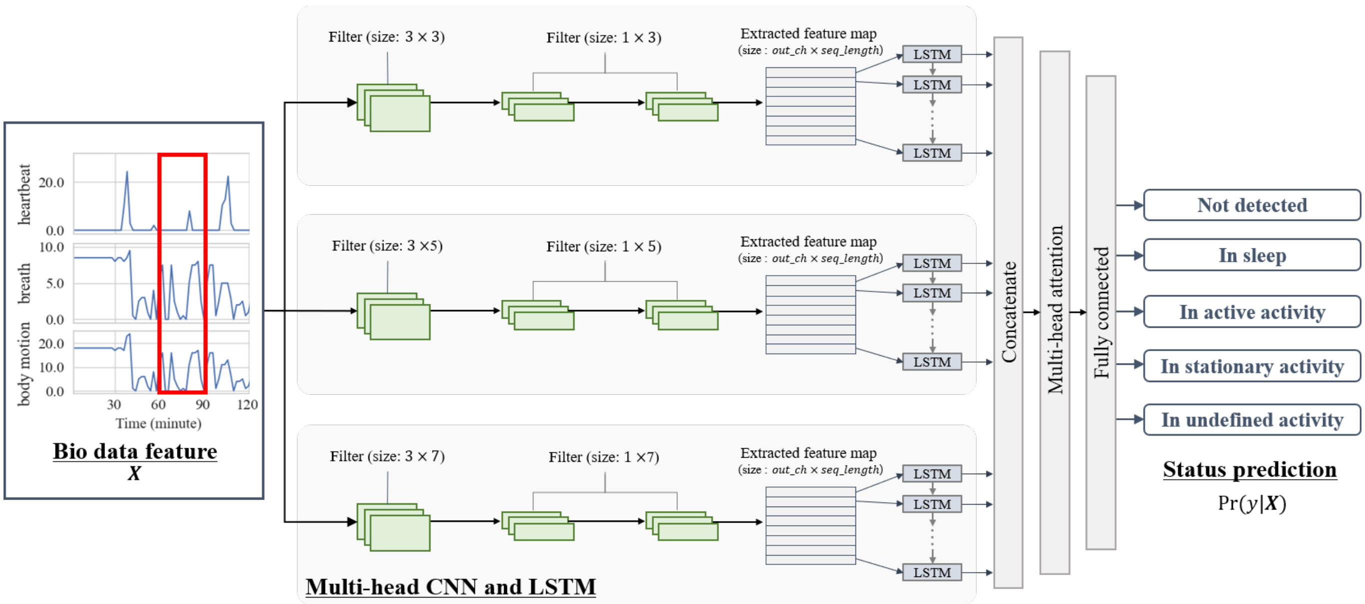

Fig. 4. Proposed deep learning-based user status estimation model with a multi-head CNN and attention mechanism

III. Status Classification from Biometric Feature

In this section, a deep learning-based status classification model is proposed that extracts and analyzes the multi-level features from biometric data. The proposed model classifies the status of a subject in one of the predefined status classes: 1) not detected, 2) in sleep, 3) in an active activity, 4) in a stationary activity, and 5) in an undefined activity. In advance of introducing the proposed classification model, we provide details on the dataset.

A. Dataset Generation and Data Preprocessing

In each living space of over 100 individual single households, a radar-based sensor with the specifications given in Table I was installed as described in Fig. 1, and the sensor measures the heart rate, respiratory rate, and body motion every 2 minutes. The biometric data were collected under uncontrolled environments at different times for each individual from 2020 to 2021. A part of the collected dataset is illustrated in Fig. 3 as an example. Only the data that are reliably collected without a network failure for a reasonable period are used for this study. In addition, whereas this study utilizes only heart rate, respiratory rate, and body motion data, the method proposed in this paper can be easily replaced with other types of biometric features.

The biometric data are labeled by experts as one of the five predefined statuses of the subject, such that the subject is 1) not detected, 2) in sleep, 3) in an active activity, 4) in a stationary activity, and 5) in an undefined activity. Min-max normalization is applied to each dimension of the biometric features. The continuous time-series data are then segmented into fully non-overlapping subsequences in the length of 30 minutes. Overlapping subsequences can be used in a practical system, but the non-overlapping subsequences are used in this paper to prevent the training, validation, and test dataset for assessment from containing partially overlapping information. The most frequently detected status in the period of the segmented subsequence is assigned as the final status of the segmented subsequence. The distribution of the generated dataset is described in Table II.

B. Status Classification with Multi-head CNN

The objective of the user status classification problem is defined as

$$\Pr(y \mid \boldsymbol{X}), \tag{1}$$

where y denotes the status class, and $\boldsymbol{X}$ is the input sequence of biometric features. That is, the status classification model aims to estimate the status of a user from the observed biometric data. Considering the biometric data $\boldsymbol{X}$ as a multivariate time series, it can be decomposed into low and high-level features like the other types of data, such as signals and images. The time-series data about human activities are often perceived as a combination of actions in low and high levels, and the extraction of features in diverse levels, therefore, can facilitate the status classification.

To extract the features from the time-series data, CNNs have been widely used in various types of research around biometric data analysis. However, vanilla CNNs have shown their limitations as they capture the features at a defined resolution. In this paper, we thereby utilize CNNs in a multi-head structure to identify the features in diverse levels from the biometric data. To additionally capture the temporal relations of the features, LSTMs are applied afterward to each of the feature maps extracted by the multi-head CNN. In addition, a multi-head self-attention mechanism is used as well to differently weigh the importance of the extracted features. The

TABLE III
CONFUSION MATRIX FOR ERROR MEASUREMENTS

Actual \ Predicted	Positive	Negative
Positive	True positive (TP)	False negative (FN)
Negative	False positive (FP)	True negative (FN)

self-attention mechanism computes attention weights for each hidden unit of the LSTM from all CNN heads to generate context vectors as a weighted sum of the hidden units.

The proposed classification model is illustrated in Fig. 4. A 30-minute long multidimensional biometric data consisting of heart rate, respiratory rate, and body motion information is fed into each head of the multi-head CNN designed with different filter sizes. Each head of the multi-head CNNs extracts the feature maps in different resolutions, and LSTMs additionally insert the information about temporal causalities into the captured features. The extracted features are differently scaled based on the attention weights computed by the multi-head attention mechanism, which is then mapped into the probability of each status class through a fully connected layer with softmax.

IV. PERFORMANCE EVALUATION

In this section, the performance of the proposed status classification model is evaluated in terms of its classification accuracy.

A. Evaluation Metric

The performance of the proposed model is evaluated in accuracy and F_1 score of the classification results, which are computed based on the confusion matrix illustrated in Table III. The accuracy and F_1 score are given by

$$\text{Accuracy} = \frac{\sum_{i=1}^{N} \mathrm{TP}_i}{\sum_{i=1}^{N} \mathrm{TP}_i + \sum_{i=1}^{N} \mathrm{FP}_i}, \tag{2}$$

and

$$\mathrm{F}_1 \text{ score} = \frac{1}{N} \sum_{i=1}^{N} \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \frac{1}{2}\left(\mathrm{FP}_i + \mathrm{FN}_i\right)}, \tag{3}$$

where TP_i, TN_i, FP_i, and FN_i are the numbers of the true positive, true negative, false positive, and false negative instances for the i-th class from N classes. The ranges of both accuracy and F_1 score are $[0, 1]$, where a value closer to 1 implies better classification performance. While accuracy is the most well-known metric for classification problems over various domains, the macro F_1 score is widely used to evaluate the performance of a classifier on a dataset with imbalanced class samples like the generated dataset.

TABLE IV
F_1 SCORE OF THE STATUS CLASSES (%)

Model	Accuracy	F_1 score
Gaussian Naive Bayes	68.0	**55.8**
k Nearest Neighbor	78.8	51.6
Support Vector Machine	82.8	52.6
Random Forest	**83.3**	54.9
Multi-layered Perceptrons (16-8)	**85.6**	**65.8**
Multi-layered Perceptrons (16-16-8)	85.3	65.8
Multi-layered Perceptrons (16-16-16-8)	85.2	62.4
Single-head CNN and LSTM	85.7	69.1
Single-head CNN and LSTM with attention	**86.7**	**71.5**
Multi-head CNN and LSTM	86.4	71.5
Multi-head CNN and LSTM with attention	**87.1**	**73.5**

B. Experimental settings and results

For performance evaluation of the proposed classification model, experiments on the methods with conventional machine learning and deep learning techniques are conducted. For conventional machine learning techniques, Gaussian naïve Bayes (NB), k-nearest neighbor (kNN), support vector machine (SVM), and random forest (RF) are applied as the benchmarks of the performance comparison. The conventional machine learning methods are implemented with the scikit-learn library in python. For baselines, multi-layered perceptions (MLP) with different numbers of layers and nodes are investigated. In addition, the models composed of CNNs followed by LSTMs with and without an attention mechanism are used for more advanced baselines. A multi-head self-attention mechanism is used for advanced baselines, and the multi-head CNN with LSTM are respectively implemented with 3 layers of CNNs with the convolutional filter sizes of (3, 5, 7) and 2 layers of bidirectional LSTMs with the hidden size of 16. For single-head CNN models, the 4 layers of CNN is used with the convolutional filter size of 5 for all layers. A fully connected layer is added before the softmax layer for all CNN and LSTM-based models. The deep learning models are implemented with TensorFlow and Keras in python. All deep learning models are trained using Adam optimizer with a learning rate 0.001 until their losses are converged with the learning rate decay and early stopping. For training, validation, and test, 60%, 20%, and 20% of instances randomly selected from the generated dataset are used, respectively. For a fair assessment, 10 independent trials of experiments are conducted, and the classification accuracy of all trials are averaged.

The experimental results are provided in Table IV. The results show that the deep learning approaches greatly outperform the conventional machine learning approaches in terms of F_1 score while there are not remarkable improvements in terms of accuracy. This implies that the deep learning approaches are more robust to the imbalanced dataset compared to the conventional machine learning approaches. In the meantime, the deep learning-based classification models tend to be easily overfitted when the fully connected layers are heavily stacked because the input feature of the biometric data is relatively simple in its contexts and small in its size compared to datasets in the other domains. However, the deep CNN and LSTM models show improved performance compared to the MLP models as the CNN and LSTM facilitate more advanced feature extraction before the last fully connected layer. In the same manner, the proposed model composed of multi-head CNN and LSTM with an attention mechanism outperforms all the other models with 2.8% to 31.7% improvements in F_1 score as it successfully extracts the contexts in diverse levels and their temporal causalities with proper importance weights. In summary, the experimental results on the empirically generated dataset show that the proposed status classification model achieves performance improvements compared to all benchmark and baseline models both in terms of accuracy and F_1 score.

V. Conclusion

In this paper, the biometric data (i.e., heart rate, respiratory rate, and motion volume) collected in the living spaces of elders under uncontrolled environments are investigated, and a classification model is proposed that estimates a user's status in one of the five predefined classes (i.e., not detected, in sleep, in an active activity, in a stationary activity, and in an undefined activity). In particular, the status classification model is designed to support the extraction of features in diverse resolutions and their temporal causalities with importance weights. The experimental results show that the proposed model enhances the status classification performance by up to 31.7% in F_1 score and 4.6% in accuracy. As the results on the empirical data collected under uncontrolled environments verify that the proposed model achieves noticeable improvements in F_1 score, the effectiveness and practicality of the proposed model are shown. In the meantime, to improve the overall classification accuracy for effective service provision, the issue of imbalanced class in the training dataset requires to be comprehensively managed in the future work of this study. In addition, to consider the scalability of the proposed status classification system, automatic labeling on the generated data also needs to be considered while handling the privacy concerns as well.

Acknowledgment

The authors of this paper thank Seung Chul Kim from KULS, João Garcia from Ubiwhere, and Ricardo Gonçalves from PROEF for co-working to develop the dataset and prototype for the proposed system as partners of a research project (KIAT, Project No. 0011879).

References

[1] N. Balta-Ozkan, R. Davidson, M. Bicket, and L. Whitmarsh, "Social barriers to the adoption of smart homes," *Energy Policy*, vol. 63, pp. 363–374, 2013.

[2] Organisation for Economic Co-operation and Development (OECD). The Future of Families to 2030. Accessed: Feb., 2020. [Online]. Available: https://doi.org/10.1787/9789264168367-en

[3] S. Majumder, E. Aghayi, M. Noferesti, H. Memarzadeh-Tehran, T. Mondal, Z. Pang, and M. J. Deen, "Smart homes for elderly healthcare—recent advances and research challenges," *Sensors*, vol. 17, no. 11, 2017.

[4] H. Mshali, T. Lemlouma, M. Moloney, and D. Magoni, "A survey on health monitoring systems for health smart homes," *International Journal of Industrial Ergonomics*, vol. 66, pp. 26–56, 2018.

[5] J. Kim, "How much do we know about the use of smartphones in the silver generation?: Determinants of the digital divide within the silver generation," *Information Society & Media*, vol. 21, no. 3, pp. 33–64, 2020.

[6] M. Bahache, J. P. Lemayian, W. Wang, and J. Hamareh, "An inclusive survey of contactless wireless sensing: A technology used for remotely monitoring vital signs has the potential to combating covid-19," 2020.

[7] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomedical Signal Processing and Control*, vol. 13, pp. 102–112, 2014.

[8] A. Ni, A. Azarang, and N. Kehtarnavaz, "A Review of Deep Learning-Based Contactless Heart Rate Measurement Methods," *Sensors*, vol. 21, no. 11, 2021.

[9] F. Zhang, C. Wu, B. Wang, M. Wu, D. Bugos, H. Zhang, and K. J. R. Liu, "Smars: Sleep monitoring via ambient radio signals," *IEEE Transactions on Mobile Computing*, vol. 20, no. 1, pp. 217–231, 2021.

[10] Y. Lee, J. Y. Park, Y. W. Choi, H. K. Park, S. H. Cho, S. H. Cho, and Y. H. Lim, "A Novel Non-contact Heart Rate Monitor Using Impulse-Radio Ultra-Wideband (IR-UWB) Radar Technology," *Scientific Reports*, vol. 8, no. 13053, 2018.

[11] C. H. Chang and W. W. Hu, "Design and implementation of an embedded cardiorespiratory monitoring system for wheelchair users," *IEEE Embedded Systems Letters*, vol. 13, no. 4, pp. 150–153, 2021.

[12] H. B. Kwon, S. H. Choi, D. Lee, D. Son, H. Yoon, M. H. Lee, Y. J. Lee, and K. S. Park, "Attention-based lstm for non-contact sleep stage classification using ir-uwb radar," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3844–3853, 2021.

[13] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "Ecg heartbeat classification: A deep transferable representation," in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 2018, pp. 443–444.

[14] N. Ahmed, A. Singh, S. KS, G. Kumar, G. Parchani, and V. Saran, "Classification Of Sleep-Wake State In A Ballistocardiogram System Based On Deep Learning," 2020.

[15] S. Purushotham, C. Meng, Z. Che, and Y. Liu, "Benchmarking deep learning models on large healthcare datasets," *Journal of Biomedical Informatics*, vol. 83, pp. 112–134, 2018.