Applications of deep learning methods in digital biomarker research using noninvasive sensing data

Hoyeon Jeong¹, Yong W Jeong¹, Yeonjae Park¹, Kise Kim², Junghwan Park³ and Dae R Kang¹,4

Abstract

Introduction: Noninvasive digital biomarkers are critical elements in digital healthcare in terms of not only the ease of measurement but also their use of raw data. In recent years, deep learning methods have been put to use to analyze these diverse heterogeneous data; these methods include representation learning for feature extraction and supervised learning for the prediction of these biomarkers. Methods: We introduce clinical cases of digital biomarkers and various deep-learning methods applied according to each data type. In addition, deep learning methods for the integrated analysis of multidimensional heterogeneous data are introduced, and the utility of these data as an integrated digital biomarker is presented. The current status of digital biomarker research is examined by surveying research cases applied to various types of data as well as modeling methods. Results: We present a future research direction for using data from heterogeneous sources together by introducing deep learning methods for dimensionality reduction and mode integration from multimodal digital biomarker studies covering related domains. The integration of multimodality has led to advances in research through the improvement of performance and complementarity between modes. Discussion: The integrative digital biomarker will be more useful for research on diseases that require data from multiple sources to be treated together. Since delicate signals from patients are not missed and the interaction effects between signals are also considered, it will be helpful for immediate detection and more accurate prediction of symptoms.

Keywords

Biomarkers, vital signs, deep learning, classification, electrocardiography, wearable electronic devices

Submission date: 6 May 2022; Acceptance date: 4 October 2022

Introduction

The United States National Institutes of Health Biomarkers Definitions Working Group defines “biological marker (biomarker): a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.”¹ Biomarkers are mainly measured from blood or tissue samples; genetic biomarkers are discovered through various medical and statistical methods. Meanwhile, the rapid development of various digital measuring instruments that record and analyze the human body has enabled their use in clinical practice. In light of this, digital biomarker research aimed at utilizing information obtained from biosensors such as biomarkers is on an expanding trajectory. The Journal of Digital Biomarkers defines a digital biomarker as follows: “Digital biomarkers are defined as objective, quantifiable physiological and behavioral data that are collected and measured by means of freely available devices”².

¹Department of Biostatistics, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea
²School of Health and Environmental Science, Korea University, Seoul, Republic of Korea
³MEZOO Co., Ltd., Wonju, Republic of Korea
⁴Department of Precision Medicine, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea

Corresponding author:
Dae R Kang, Department of Precision Medicine and Biostatistics, Yonsei University Wonju College of Medicine, 20, Ilsan-ro, Wonju, Gwanwon-do 26426, Republic of Korea.
Email: dr.kang@yonsei.ac.kr

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).
of digital devices such as portables, wearables, implantables, or digestibles. The data collected are typically used to explain, influence, and/or predict health-related outcomes.” (Dorsey ER, editor. Digital biomarkers. Basel: Karger.) It has been reported that digital biomarkers tend to be non-invasive or less invasive than conventional biomarkers, and are beneficial in terms of their ease of measurement and cost reduction. Although traditional statistical methods are still used to find digital biomarkers, many machine learning and deep learning methods have been proposed because of the necessity of dealing with raw data of various modalities as they are.

Several cases of the clinical application of digital biomarker research methods for a range of diseases have been surveyed. The use of digital biomarkers collected through mobile and wearable devices in a personalized, non-invasive and continuous way for diagnosis of Alzheimer’s disease has been investigated. The heart rate variability (HRV) as a digital biomarker for continuous telemedicine in daily life through wearable devices has been investigated. Moreover, surveys have also been conducted on applying artificial intelligence (AI) technology to digital biomarker research. Studies using electrocardiogram (ECG) signals as digital biomarkers extracted through machine-learning methods have been investigated for detecting atrial fibrillation (AF). Studies focusing on the application of AI technology on ECG signals for cardiovascular disease management have been surveyed. The deep learning method, one of the AI technologies, has the advantage of pre-processing data by itself without complicated processing by humans and the advantage of high model accuracy. Biosignal processing and analysis using machine learning and deep learning methods have been investigated. On the other hand, in a different aspect, deep learning methods for the fusion or representation of multimodal data, rather than a model for a single modality, have been surveyed.

In this review, we introduce several kinds of digital biomarkers used in recent times, report their clinical applications, and review deep learning techniques used for the supervised learning and dimensionality reduction. As deep learning methods, supervised learning is suitable for predicting a biomarker, and dimensionality reduction through representation learning is useful for processing with multiple biomarkers at once. We first present examples of the application of deep learning in biomarker research, and then comprehensively review research cases of representation learning that encompasses adjacent fields, and present this as a future research direction for integrated biomarker research for multimodal data. It mainly focuses on sending data obtained through noninvasive methods and excludes invasive methods or image biomarkers. We mainly used PubMed to search for clinical cases of digital biomarkers and Google Scholar for technical reporting of deep learning. The search results of RISS were used for research cases written in Korean.

**Clinical use cases of digital biomarkers**

Different types of sensors are used alone or in combination for obtaining digital biomarkers for a variety of clinical applications. They are largely divided into electrode-based sensors and inertia-based sensors. ECGs are representative electrode-based sensors. The ECG sensor is attached as a patch to the skin surface around the heart in wired or wireless form. Inertia-based sensors are sometimes used in the form of patches or wearable devices, but they are also embedded in smart devices such as smartphones. In this section, clinical cases using each type of sensor, alone or in combination, are introduced, and Table 1 provides a brief summary.

**ECG signals**

For ECG signals, it is possible to extract derivative signals such as heart rate (HR), HRV, and RR interval, which can also be used together or separately. Many previous studies have considered ECG-based data as digital biomarkers. In the past, cardiac rehabilitation response was evaluated using a wearable ECG sensor, and amyotrophic lateral sclerosis (ALS) was monitored by measuring HRV in a disposable electrode patch-based ECG sensor, and assessed by analyzing HRV and respiration through ECG devices in a study the health statuses of fishermen.

**Activity data**

Motion-based digital biomarkers have also been proposed, which are usually measured using accelerometers and gyroscope sensors. Several use cases of motion-based digital biomarkers have been reported, particularly in Parkinson’s disease (PD), where movement or mobility has been evaluated as a digital biomarker. The symptoms of PD were evaluated by measuring tremor and bradykinesia, and the gait fluctuations in PD patients were reported using magnetometers and barometers, as well as accelerometers and gyroscopes. Monitoring neuromuscular disorders in patients with multiple sclerosis and PD using a wearable magnetoinertial sensor has been reported. Wearable inertial sensors have also been used to improve sternal precautions. A study involving patients with cardiovascular disease and focused on the association between the movement of a wrist-worn fitness tracker and existing biomarkers (blood and body composition) has been reported.

**Other single modalities**

In addition to the major modalities mentioned above, sensor data from various modalities are used as digital biomarkers. HR can also be extracted from a photoplethysmogram (PPG) as an optical method, a smartphone-based PPG used to detect diabetes with a deep neural network...


| Modality                  | Clinical use cases                                                                 |
|---------------------------|-----------------------------------------------------------------------------------|
| ECG signals               | Cardiac rehabilitation response,10 ALS,11 health statuses of fishermen12          |
| Activity data from inertial sensors | PD,13-16 sternal precautions,17 cardiovascular disease18                      |
| Other single modalities   | PPG for diabetes,19 EEG for mild cognitive impairment,20 Sound for asthma,21 skin-adhesive patches for monitoring of sweat electrolytes,22 pH for periodontal disease and nitric oxide concentration biomarker for heart failure using wearable biochemical sensors,23 fluorescence image from a cellulose-based wearable patch for an analysis of sweat components,24 |
| Multimodal data           | ECG and PPG for asymptomatic AF;25 EEG, ECG, speech patterns, and kinematics for PD;26 ECG, respiratory rate, fluid status, physical activity, and posture for heart failure;27 ECG, thermistor, and accelerometer for stress response;28 multimodal wearable sensors for measurement of gait and voice;29 accelerometer, barometer, GSR, temperature and photo sensors for evaluation of fatigue;30 smartphone records (call logs, text messages, and accelerometers) for social anxiety;31 accelerometer data from the smartwatch for PD;32 activity and sleep data for gastroenteritis;33 multimodal ambient sensors for heart failure decomposition34 |

ECG: electrocardiogram; ALS: amyotrophic lateral sclerosis; PD: Parkinson's disease; PPG: photoplethysmogram; EEG: electroencephalogram; AF: atrial fibrillation; GSR: galvanic skin response.

A wearable electroencephalogram (EEG) is also being used as a digital biomarker to identify mild cognitive impairment.20 Sound can also be used as a digital biomaker, as evidenced from existing studies focusing on recording coughing sounds using a smartphone as a digital biomarker for asthma.21 Furthermore, using skin-adhesive patches have also been reported. Sweat and skin body temperature were measured using a skin-adhesive electrode-based radio frequency identification bandage,22 and pH (biomarker for periodontal disease) and nitric oxide concentration (biomarker for heart failure) were measured using wearable biochemical sensors.23 Smartphones have also been used for detecting biochemical reactions; smartphones were used in a past study to capture the fluorescence image from a cellulose-based wearable patch, which was then used as a biomarker for analysis of sweat components.24

**Multimodal data**

A number of studies have used multimodal data from heterogeneous sensors. As a clinical case in which ECG signals and other multimodal signals are used for cardiovascular disease, the risk of asymptomatic AF has been screened using an ECG patch and a wrist-worn PPG device.25 Furthermore, many physical and cognitive studies of PD have been conducted using EEG, ECG, speech patterns, and kinematics.26 ECG and ECG-derived parameters, respiratory rate, fluid status, physical activity, and posture have been used to monitor patients with heart failure.27 The stress response of music performers was also analyzed using an ECG sensor, thermistor, and accelerometer.28 A study using a multimodal wearable sensor has to measure gait and voice also been reported.29

Further, dimensionality of multimodal data obtained from an accelerometer, barometer, galvanic skin response (GSR) electrode, temperature sensors, and photo sensors of multisensor wearable devices was reduced using an unsupervised casual convolutional neural network (cCNN), and fatigue was evaluated.30 Multimodal data from smart devices have also been used as digital biomarkers. For example, correlation analysis of the severity of social anxiety using smartphone records (call logs, text messages, and accelerometers) has been reported,31 accelerometer data (activity, tremor, and movement) from the smartwatch of a PD patient was collected and stored,32 and the severity of symptoms in gastroenteritis patients was estimated through activity and sleep data from smartphones and smart watches.33 Furthermore, ambient sensors placed at a distance from the body have also been used as essential digital biomarkers. For example, in a previous study, the multimodal sensor data, which were used as digital biomarkers for early signs of heart failure decomposition, such as motion sensor, HR, respiration rate, and toss-and-turn counts were transmitted to the cloud, and their correlation with the administered questionnaire was analyzed.34

**Notable deep learning models**

The accumulation of large volumes of data and the exploitation of parallel computing resources have accelerated the development of deep learning technology; over the past decade, these deep learning models have been used in various fields of academia and industry, and they are also being actively applied in medicine and healthcare. In this section, we discuss some of the
notable deep learning methods and their derivatives and review their effectiveness.

**Deep neural networks**

Deep learning is a method of building and training a DNN with hidden layers to be mainly used for inference and prediction. Deep learning can be broadly divided into supervised and unsupervised learning. It can be further subdivided according to the type of hidden layer used. Unless otherwise specified, deep learning methods use a fully connected (FC) layer that connects all nodes, and the stacked structure is called a multilayer perceptron (MLP). More detailed techniques based on the types of hidden layers are discussed in the next subsections. Hidden layers are stacked, and the output layer is usually used for classification, which is the case in binary and multiclass classification, here, the sigmoid function and the softmax function are used as activation functions. Figure 1 shows a schematic diagram of this structure. Deep learning models have the advantage of improved model performance through spatial distortion and weight transfer and update through deep structures, which means they include pre-processing, such as feature extraction and dimensionality reduction internally. Due to these properties, digital signals that have undergone minimal pre-processing can be used as digital biomarkers through deep neural networks.

**Convolutional neural networks and graph neural networks**

The remarkable performance of AlexNet—an image classifier based on a convolutional neural network (CNN)—which won the ImageNet challenge in 2012 for image classification with an overwhelming margin, greatly improved the wider application of deep learning methods. Convolution automatically enabled feature extraction, which had previously been done manually. Research on stacking hidden layers to increase accuracy has been developed; and as one of them, the inception model, which reduces the number of parameters by concatenating a small-sized convolution operation, has been proposed. ResNet, which transmits the original value by adding a residual block as a skip connection, was also proposed. In contrast, models that apply neural networks to graph structures as connections between nodes have also shown remarkable performance; representative examples include the graph neural network (GNN) and graph convolutional network (GCN).

**Recurrent neural networks**

Meanwhile, another notable study proposed a deep neural network model based on natural language processing. This model is a recurrent neural network (RNN), which is suitable for dealing with sequential information such as text or time series data. In addition, the following methods were proposed: an embedding method into a low-dimensional vector (Words to vectors (Word2vec)), a method for transmitting information in long sentences (long short-term memory, LSTM), its lightweight model (gated recurrent unit, GRU), a method for paying attention to time points, and a self-attention method without sequential recurrence.

**Representation learning and transfer learning**

Representation learning, also known as feature learning, is used in unsupervised learning to extract features. Convolutional operation of the CNN model is also a type of representation learning, but other DNN-based methods are also widely used. The restricted Boltzmann machine, which is a bipartite graph with latent variables, and the deep Boltzmann machine (DBM), with hidden layers inside, are representative examples. An autoencoder with an encoder–decoder structure and a stacked autoencoder (SAE) with a hidden layer added have also been used.
Additionally, in terms of reusing model weights, transfer learning, which learns a model with big data from one domain and uses it in other domains, is sometimes used.

Deep learning methods for prediction and dimensionality reduction in noninvasive sensing data

In this section, we discuss the application of deep learning methods to noninvasive sensing data in digital biomarker research, which is the subject of this review article. We considered the two main aspects of supervised learning and representation learning: prediction (including classification and regression) and dimensionality reduction (including feature extraction and decomposition). Figure 2 compares the conceptual difference between Figure 2(a) discovering digital biomarkers through supervised learning for each modality and Figure 2(b) discovering an integrated digital biomarker through supervised learning after integrating multiple modalities through representation learning. Table 2 compares the CNN, GNN, and RNN methods, which are mainly used in the supervised learning of digital biomarker research.

The following paragraphs are divided according to the type of sensor, as in the clinical case study in the previous section. ECG signals, activity data from inertial sensors, and other modalities are described, and integrated analysis cases and integrated analysis of biosignals accompanying environmental information are also covered. Table 3 lists the application cases of the deep neural network models according to the type of modality for each sensor mentioned.

**ECG signals**

The CNN method is applied to the ECG signals in the following cases. Prediction accuracy of early coronary artery occlusion by constructing a CNN for ECG signals exceeded that by experts. In one study, mental stress was measured with high accuracy of 87.39% by using a CNN model on ultrashort ECG signals rather than the existing HRV-based method (71.17%); for HRV, extracted parameters were used for a length of 5 min, and for ECG signals, a segment with a length of 0.8 s was used. A CNN analysis of the ECG signals was used for noninvasive assessment of dofetilide plasma concentration, which is a nonlinear regression model for ECG segments. Later, these researchers also applied a CNN model for the identification of AF during sinus rhythms of ECG signals. Another group also classified AF by applying the CNN model to

| Table 2. Supervised learning method for discovering digital biomarkers. |
|---------------------------------|---------------------------------|---------------------------------|
| **Deep Learning method** | **Main role** | **Structure and learning method** |
| CNN | Extracting feature maps from multidimensional data | A convolutional layer is stacked, an FC layer is connected, and a skip connection is sometimes added. Minimizes cross-entropy loss between actual and predicted values |
| GNN | Utilize not only features of entities but also link information among entities | For graph-structured data with features and links, stacking convolution operations (GCN), etc., and minimizing loss, thus updating features |
| RNN | Prediction of multidimensional sequence data | Used as an LSTM or a GRU model that preserves weights information. It is mainly used to extract features from sequence data through a CNN layer and connects LSTM layers |

**Figure 2.** Utilization of multimodal data as digital biomarkers. Here, (a) is for each modality, whereas (b) is for multimodal fusion, which can improve accuracy through representation learning.
ECG signals; furthermore, these researchers used this method to classify sleep apnea. A CNN model was applied to the evaluation of the sleep stages using instantaneous HR extracted from ECG signals. Here, two CNN layers were applied: convolution for extracting local features and a dilated convolutional block for extracting long-range features. A CNN model was used to examine changes in ischemic ST from ECG signal recordings in gait, which was then retrained via transfer learning on an Inception V3 model pretrained with ImageNet data. In another study, multiple classes of myocardial infarction were classified using a model combining CNN and LSTM layers for ECG signals from a portable device. Moreover, a study identified mental stress by combining CNN and LSTM layers for ECG and respiration signals. Here, the CNN and LSTM layers were mixed for each modality, and they were then concatenated to connect the FC layer and the sigmoid layer. It showed a high prediction performance (area under curve (AUC), 0.92) compared with other machine learning methods (0.63–0.80).

Activity data
Studies applying the CNN method to data from accelerometers and gyroscopes have also been reported. PD was diagnosed using the CNN method from smartphone-based accelerometer data, and the severity of PD was predicted by classifying the accelerometer and gyroscope data using a CNN model. Furthermore, a study classified the respiratory patterns using a CNN model on accelerometer and gyroscope data. For motion data from the accelerometer of a smartwatch, two convolution layers and a softmax layer were constructed to classify the five rehabilitation actions. One study predicted freezing of gait by applying an LSTM model to wearable accelerometer data of PD patients, and the accuracy was improved through transfer learning by splitting the sample. A study on detecting abnormal behavior by applying autoencoder and LSTM models to the movement data of dementia patients has been reported. Abnormal behaviors were extracted with an autoencoder that learned the patterns of normal behaviors, after which they were labeled and learnt by the LSTM model. To diagnose PD and essential tremor, high-resolution tremors from patients’ smartphones were captured and classified by combining CNN and LSTM layers. There are also studies wherein a GCN model was applied to human activity recognition (HAR) of time-series data from inertial sensors of smartphones. By dividing the samples, each sample was composed of a graph, and the temporal connection was defined as an undirected

| Modality                                      | CNN | CNN and LSTM | LSTM | GNN | Other models                      |
|-----------------------------------------------|-----|--------------|------|-----|-----------------------------------|
| ECG signals (including HR and HRV)            | Brisk et al.50, Hwang et al.51, Attia et al.52,53, Erdenebayar et al.54, Urtnasan et al.55, Sridhar et al.56, Xiao et al.57 | Lui and Chow58, Seo et al.59 |
| Activity data from inertial sensors (including accelerometer and gyroscope) | Zhang et al.60, Sieberts et al.61, McClure et al.62, Lee et al.63 | Varghese et al.64, Torvi et al.65 | Mondal et al.66 | Autoencoder and LSTM67 |
| Other single modalities (including BCG, PSG, audio sounds, GPS, EEG) | Ahmed et al.68, Kim et al.69, Srivastava et al.70, Ahmed et al.71 | Bidia72, Zhang et al.73 | MLP74, Encoder75 |
| Integration of multimodal data                | Jung et al.76 | Rutkowski et al.77, Han et al.78 | Dong et al.79 | CNN and Encoder30, Transformer40, DNN (not specified)81 |
| Biosignals accompanying environmental information | Kanjo et al.82, Mou et al.83 | He et al.84 |

ECG: electrocardiogram; BCG: ballistocardiography; CNN: convolutional neural network; ECG: electrocardiogram; EEG: electroencephalogram; DNN: deep neural network; GNN: graph neural network; GPS: global positioning system; HR: heart rate, HRV: heart rate variability, LSTM: long short-term memory; MLP: multilayer perceptron; PSG: polysomnography.
edge. After constructing the two GCN layers, an FC layer of softmax activation was combined. In terms of accuracy, the results are higher (100%) than those of existing methods (89.9%–99.9%).

**Other single modalities**

In addition, data from other single modalities with DNNs have also been analyzed. Sleep-wake states were classified by constructing CNNs on ballistocardiography (BCG) signals. Compared with polysomnography (PSG), this was a simpler and noncontact procedure but showed high accuracy of 94.90%. Furthermore, developmental disabilities were predicted by analyzing smartphone touch information using a CNN model. A CNN model has also been used to diagnose chronic obstructive pulmonary disease by applying it to breathing sounds. A study inferred lung anomalies by applying a CNN model to audio and inertial sensor data from smartphones for pulmonary patients. Depression was diagnosed by analyzing voice samples of PD patients with MLP, which showed high prediction accuracy of 77% compared with other prediction models based on machine learning (0.62–0.76). Encoders, meanwhile, have been applied to predict morbidity and mortality. To measure biological age acceleration, the gait data from the wearable sensor was embedded into the encoder, and age was added to define the biological age. Comparisons between motion-based aging and blood-based aging were found to be related. Depression has also been predicted by applying a GNN model to data from smartphones and wearable devices. The survey results, sensing data (activity, sleep, and HR) and global positioning system (GPS) data were used as node features, and clinical data (diagnosis) were used as true labels. For edges between nodes, similarity etc., was used. Comparing the prediction values from the GNN model and true labels indicated that the accuracy score was higher (80% or more) than those of the traditional method (70% or less). Another study detected driver fatigue by connecting EEG channels with a GCN model. An adjacency matrix was constructed between channels with partial directed coherence as a causal analysis. Local area information and global connection information were combined with the GCN model, and classified as softmax through FC. It showed high accuracy of 83.84% compared with other feature-based deep learning models (63.69%–83.90%).

**Integration of multimodal data**

In digital biomarker research, deep learning methods have been reported for the integration of multidimensional data. There was also a case of detecting emergency events using a deep learning model that combines video, audio, activity, and dust sensor data. The data correspond to video from CCTV in the home, voice from an AI speaker, activity from a smartphone accelerometer, and dust from the air purifier. As a method for image object detection, a region-based CNN model pretrained with Inception V2 was used, and a Gaussian mixture model was used for sound classification. The fusion method showed a higher accuracy of 94% than many of the models for each modality (89%–94%). The cCNN model has been used as a dimensionality reduction method to evaluate fatigue level (obtained through questionnaire) on multimodal data from the accelerometer, PPG, temperature, GSR, barometer, etc., in which an encoder based on causal dilated convolutions and a triplet loss inspired by Word2vec are combined for representation learning using multiple sensor data. In a study to investigate the brain correlation between task-load and dementia, for classification of the dementia stage from EEG signals and event-related potential with task-load stimuli, tensor-train-decomposition-based dimensionality reduction and an LSTM-based classifier were used. Another study classified a pilot’s mental stress by applying a CNN model to EEG signals and LSTM models to ECG, respiration, and electrodermal activity (EDA) signals. Signals were concatenated in the FC layer and classified as softmax. The stimuli and the corresponding labels were predefined. The approach showed a higher accuracy of 85.2% than the EEG monomodal (77.7%) or other machine learning methods (55.6%–81.4%). One study predicted behavior by applying a transformer (multhead self-attention) model to data from wrist-worn devices (step count, HR, and sleep status) and daily survey data. The model compresses the dimensions with a convolutional encoder, passes through a 4-headed attention layer, and connects a classifier. It showed a higher area-under-curve value of 94% compared with the CNN model (61%) and the eXtreme gradient boosting (XGBoost) model (74%). In addition, the performance improvement was confirmed by dividing the data and using the pretrained model for fine-tuning the small-scale data. In another study, influenza-like symptoms were recognized by applying a GNN model to multimodal mobile sensing data. Data used here are GPS trajectory, Bluetooth encounter trace, and activity (accelerometer and gyroscope), and self-reports are used to infer at least one influenza-like symptom. For each sensing data, features were extracted through the GNN model using a GraphSAGE convolutional layer and then concatenated and classified through FC with sigmoid. It showed higher AUC of 95.39% than other GNN models (67.25%–91.48%) as well as existing machine learning models (65.32%–74.43%). A case of predicting metabolic syndrome by merging lifelog data from wearable devices as well as clinical data and genetic data using unspecified DNNs has been reported.

**Biosignals accompanying environmental information**

In addition, there have been reports of cases of designing a deep learning architecture that also considers biosignals accompanying environmental information. For ECG,
EDA, audio, and video data, the prediction of emotional dimension by fusion of the multimodality with bidirectional LSTM model using Pearson’s correlation coefficient as an objective function has been reported. The LSTM layers for each modality were constructed and integrated into the LSTM layer, which showed better results than the single-modality models. In another case, emotions were classified using a model that combines CNN and LSTM layers for not only body physiology data but also environmental information from mobile sensors, which showed a higher accuracy of 94.7% than unimodal-based methods (57.4%–87.4%). On-body physiological data include HR, GSR, motion, body temperature, and environmental information, including noise level, UV, atmospheric pressure, and location information. The CNN-LSTM models were constructed for each on-body, environment, and spatial information, and fused in an FC layer. In another case, by combining CNN and LSTM layers, the driver’s stress level was assessed through multimodal fusion of attention-based CNN-LSTM models using eye data, vehicle data, and environmental data. Self-attention was applied to weighting the fusion layer. Here, eye data included pupil diameter, gaze dispersion, and blinking frequency; vehicle dynamics data included steering wheel angle, brake, and acceleration; and environmental data included distance from the preceding vehicle, time of day, road conditions (lane width and number of lanes), and weather conditions (fog, sun, and rain). The labels were obtained through a questionnaire. The multimodality model showed a higher accuracy of 95.5% than the unimodality model (52.6–85.1%).

Deep learning methods for dimensionality reduction or fusion of multimodal data in other related domains

Because few studies in digital biomarker research have applied deep learning-based dimensionality reduction to multimodal data, we included studies from other related domains that applied dimensionality reduction or fusion models based on deep learning to multimodal digital data in this review. We divided these models into DBM, SAE, and deep canonical correlation analysis (DCCA) models. A comparison of the methods is presented in Table 4.

### Deep Boltzmann machine

The RBM models were used for stream integration of multiple sensors such as accelerometers and angular velocity sensors from smartphones, which consists of a hidden layer for each modality and a combined layer, achieving higher performance compared with other machine learning models. For the diagnosis of Alzheimer’s disease and mild cognitive impairment (MCI), features are extracted through DBM-based representation learning method from multimodal data of magnetic resonance imaging and positron emission tomography. Here, a deep Boltzmann machine consisting of a layer for each modality and an upper shared hidden layer is proposed, and the obtained features are classified using a support vector machine (SVM). Voice and face images from mobile phones were combined using a joint deep Boltzmann machine and used for personal identification. After learning the DBM for each unimodality and learning the shared layer, fine-tuning was performed. SVM was used for classification. In a previous study, a DBM model was used in the bimodal integration of video and text for user recommendation; the image is represented through a CNN model, whereas the text is represented through Word2vec and jointly represented in the DBM layer. In the case of the image, it was reported that more detailed features were added in the case of a single modal, and that the text prediction showed higher similarity.

### Stacked autoencoder

First, SAE models were used for integration between multimodalities. For the classification of chronic kidney disease,
a method for reducing the dimension of multimodal data such as demographic information and blood test data using an SAE model and classifying it as a softmax layer has been reported. Furthermore, a model for classifying emotions by compression using an SAE model for EEG and electromyography (EMG) data has been reported. Subjects were asked to measure the EEG and EMG signals while watching a video and to evaluate their emotions in response to the video. Pretraining for each unimodality was performed; moreover, fine-tuning was performed by combining multiple modes, and softmax was used for classification.

The following are reported cases in which SAE models are used for the mutual reconstruction of multimodal data. A study detected fatigue while driving by applying an RBM-based deep autoencoder to EEG signals and electrooculogram (EOG) signals. After constructing and joining the RBM layers for each modality, it was reconstructed. The level of fatigue was labeled by tracking eye movements, and correlation analysis using the deep autoencoder showed better results compared with unimodality. Meanwhile, a previous study to classify semantic concepts by constructing a stacked contractive autoencoder for audio, video, and text data was conducted. It involved single-modal pretraining and multimodal fine-tuning and connects the softmax layer to the classifier. Furthermore, another study predicted the missing perceptual information by constructing a multichannel autoencoder related to the three modalities of word, video, and sound. Each modality is embedded through Word2vec, associates in latent space through an SAE model, and decoded according to each modality. According to a study, by fusing multimodal data of temperature, humidity, and illuminance into an SAE model, it is possible to predict other modalities without the use of measurement records. Here, the encoder for each modality is constructed, and after integration in the latent variable layer, it is decoded into each modality again. For emotion detection, a multimodal autoencoder is applied to multimodal data of posture data and EDA data, and classification accuracy is improved by imputation through reconstruction from inter-related contexts. The missing data were input into the encoder and the complete record was obtained as the output. The decoder outputs the imputed values and uses them for the reconstruction. For the classifier pipeline, two feature vectors are concatenated and dimensionally was then reduced using PCA and classified using SVM. It was reported that the multimodal data imputation framework improved the performance of emotional recognition for multimodal data.

**Deep canonical correlation analysis**

In another reported case, emotions were recognized through DCCA for multimodal signals such as EEG and eye movement. A deep neural network was constructed for each modality, and was then fused through the CCA model and classified as an SVM model. It showed high accuracy of 94.58% compared with the existing methods (81.71%–93.97%). A different study reported multimodality fusion based on the attention mechanism for emotion recognition. For the three modalities of EEG, eye images, and eye movement, the features were extracted with pretrained ResNet with ImageNet data, and a hidden layer was constructed for each, after which they were classified as softmax by fusion with the attention method. This attention method showed a higher accuracy of 82.11% compared with the average weighted sum method (74.66%). For video and motion sensors or video and audio datasets, a temporal fusion model—the correlational RNN—was used for continuous correlation over time. The correlation between one modality and another modality for each time point was obtained using the GRU model; an encoder was constructed by connecting it with the RNN, and a corresponding decoder was constructed. It showed higher accuracy of 96.11% compared with the autoencoder model (94.4%) or the RBM model (95.8%).

**Discussion**

Deep learning-based methods as the most recent methods for discovering digital biomarkers. In addition to the widely used electrode-based and inertia-based sensors, biochemical-based sensors, satellite-based geolocation, and lifelogs from smartphones such as voice, text, and touch are indispensable digital biomarkers. Furthermore, deep learning methods are not limited to typical data characteristics or the limited functionality of prediction models. The latest methods that utilize various techniques have been proposed, and various application cases of the same have been presented.

Our survey did not collect all clinical cases related to digital biomarkers, but considers the current academic research trends related to the discovery of digital biomarkers and suggests a future research direction. In particular, the link between measurements should not be overlooked. Moreover, connections between devices or between individuals provide mutually enriched information as well as their characteristics. Meanwhile, the integration of multimodality in other related domains has resulted in tremendous progress in research through complementarity and performance improvement between modalities. However, attempts to integrate multimodality in digital biomarker research are still rare. Correlations and predictions between modalities, as well as multidimensional reduction and integrated prediction, are important topics in the selection and discovery of digital biomarkers, and further studies in this regard are needed. For such integration, preprocessing such as time synchronization or matching the format and period of the input is required. In addition, it is worth considering an additional model that prevents communication failure for each
input signal of the integrated model and reconstructs data loss in case of data loss.

The integrated digital biomarker will be more useful in studies such as neurological diseases that require data from various sources to be treated together. By considering not only traditional clinical and laboratory data, but also electrode-based and motion-based sensing data, and mobile records, it will be possible to take into account the interaction effect between signals without missing delicate signals from patients. This could help with immediate detection and more accurate prediction of symptoms. In addition, the integrated digital biomarker can be utilized in the study of environmental diseases in which not only the signal derived from the subject but also the surrounding environmental factors must be considered. It will be possible to detect changes in bio-signals for individuals exposed to environmental harmful factors, and lead to immediate action on the aspects of these symptoms.

Abbreviations

| Abbreviation | Description |
|--------------|-------------|
| AF           | atrial fibrillation |
| BCG          | ballistocardiography |
| CNN          | convolutional neural network |
| DBM          | deep Boltzmann machine |
| DCCA         | deep canonical correlation analysis |
| DNN          | deep neural network |
| ECG          | electrocardiogram |
| EDA          | electrodermal activity |
| EEG          | electroencephalogram |
| EMG          | electromyogram |
| FC           | fully-connected |
| GCN          | graph convolutional network |
| GNN          | graph neural network |
| GPS          | global positioning system |
| GRU          | gated recurrent unit |
| GSR          | galvanic skin response |
| HR           | heart rate |
| HRV          | heart rate variability |
| LSTM         | long short-term memory |
| MLP          | multilayer perceptron |
| PD           | Parkinson’s disease |
| PPG          | photoplethysmogram |
| PSG          | polysomnography |
| RBM          | restricted Boltzmann machine |
| ResNet       | residual network |
| SAE          | stacked autoencoder |
| SVM          | support vector machine |
| Word2vec     | words to vectors |

Contributorship: Conceptualization was handled by H Jeong; acquisition of data was taken care by H Jeong, YW Jeong, Y Park, K Kim; funding acquisition was done by DR Kang; investigation was carried out by H Jeong, YW Jeong, Y Park, K Kim; project administration was done by DR Kang; Resources: J Park; supervision was done by DR Kang; visualization was handled by H Jeong; Writing—original draft, Writing—review and editing, approval of final manuscript was handled by all authors.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: This study was reviewed by the Institutional Review Board of Yonsei University Wonju Severance Christian Hospital (Approval number: CR321327).

Funding: The authors received the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Korea Environment Industry & Technology Institute (KEITI) through the Development of a Personalized Service Model for Management of Exposure to Environmental Risk Factors among Vulnerable and Susceptible Individuals Program, funded by the Korea Ministry of Environment (MOE)(2021003340003).

ORCID iD: Hoyeon Jeong https://orcid.org/0000-0001-6812-9343
Dae R Kang https://orcid.org/0000-0002-8792-9730

References

1. Group BDW, Atkinson Jr AJ, Colburn WA et al. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001; 69: 89–95.
2. Babrak LM, Menetski J, Rebbin M et al. Traditional and digital biomarkers: two worlds apart? Digit Biomark 2019; 3: 92–102.
3. Kourtis LC, Regele OB, Wright JM et al. Digital biomarkers for alzheimer’s disease: the mobile/wearable devices opportunity. NPJ Digit Med 2019; 2: 1–9.
4. Owens AP. The role of heart rate variability in the future of remote digital biomarkers. Front Neurol 2020; 14: 582145.
5. Wesselius FJ, van Schie MS, De Groot NMS et al. Digital biomarkers and algorithms for detection of atrial fibrillation using surface electrocardiograms: A systematic review. Comput Biol Med 2021; 133: 104404.
6. Siontis KC, Noseworthy PA, Attia ZI et al. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nat Rev Cardiol 2021; 18: 465–478.
7. Yoon D, Jang JH, Choi BJ et al. Discovering hidden information in biosignals from patients using artificial intelligence. Korean J Anesthesiol 2020; 73: 275–284.
8. Gao J, Li P, Chen Z et al. A survey on deep learning for multimodal data fusion. Neural Comput 2020; 32: 829–864.
9. Guo WZ, Wang JW and Wang SP. Deep multimodal representation learning: A survey. IEEE Access 2019; 7: 63373–63394.
10. De Cannière H, Smeets CJ, Schouteten M et al. Using biosensors and digital biomarkers to assess response to cardiac rehabilitation: observational study. J Med Internet Res 2020; 22: e17326.
11. Garcia-Gancedo L, Kelly ML, Lavrov A et al. Objectively monitoring amyotrophic lateral sclerosis patient symptoms during clinical trials with sensors: Observational study. JMIR Mhealth Uhealth 2019; 7: e13433.

12. Wilbur RE, Griffin JS, Sorensen M et al. Establishing digital biomarkers for occupational health assessment in commercial salmon fishermen: Protocol for a mixed-methods study. JMIR Res Protoc 2018; 7: e10215.

13. Shah VV, McNames J, Mancini M et al. Digital biomarkers of mobility in parkinson’s disease during daily living. J Parkinsons Dis 2020; 10: 1099–1111.

14. Mahadevan N, Demaneule C, Zhang H et al. Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device. Npj Digital Medicine 2020; 3: 1–12.

15. Evers LJW, Raykov YP, Krijthe JH et al. Real-life gait performance as a digital biomarker for motor fluctuations: The parkinson@home validation study. J Med Internet Res 2020; 22: e19068.

16. Lilien C, Gasnier E, Gidaro T et al. Home-based monitor for gait and activity analysis. J Vis Exp 2019: 150: e59668.

17. Wang C, Goel R, Noun M et al. Wearable sensor-based digital biomarker to estimate chest expansion during sit-to-stand transitions-a practical tool to improve sternal precautions in patients undergoing median sternotomy. IEEE Trans Neural Syst Rehabil Eng 2020; 28: 165–173.

18. Rykov Y, Thach TQ, Dunleavy G et al. Activity tracker-based metrics as digital markers of cardiometabolic health in working adults: Cross-sectional study. JMIR Mhealth Uhealth 2020; 8: e16409.

19. Avram R, Olgin JE, Kuhar P et al. A digital biomarker of diabetes from smartphone-based vascular signals. Nat Med 2020; 26: 1576–1582.

20. Iliadou P, Paliokas I, Zygouris S et al. A comparison of traditional and serious game-based digital markers of cognition in older adults with mild cognitive impairment and healthy controls. J Alzheimers Dis 2021; 79: 1747–1759.

21. Rassouli F, Tinschert P, Barata F et al. Characteristics of asthma-related nocturnal cough: A potential new digital biomarker. J Asthma Allergy 2020; 13: 649–657.

22. Rose DP, Ratterman ME, Griffin DK et al. Adhesive rfid sensor patch for monitoring of sweat electrolytes. IEEE Trans Biomed Eng 2015; 62: 1457–1465.

23. Pataranuratporn P, Jain A, Johnson CM et al. Wearable lab on body: combining sensing of biochemical and digital markers in a wearable device. In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE. ISBN 1538613115, pp. 3327–3332.

24. Ardalan S, Hosseinifard M, Vosough M et al. Towards smart personalized perspiration analysis: An iot-integrated cellulose-based microfluidic wearable patch for smartphone fluorometric multi-sensing of sweat biomarkers. Biosens Bioelectron 2020; 168: 112450.

25. Galarnyk M, Quer G, McLaughlin K et al. Usability of a wrist-worn smartwatch in a direct-to-participant randomized pragmatic clinical trial. Digit Biomark 2019; 3: 176–184.

26. Ryu J, Vero J, Dobkin RD et al. Dynamic digital biomarkers of motor and cognitive function in parkinson’s disease. J Vis Exp 2019; 149: e59827.

27. Kramer F, Butler J, Shah SJ et al. Real-life multimarker monitoring in patients with heart failure: Continuous remote monitoring of mobility and patient-reported outcomes as digital end points in future heart-failure trials. Digital biomarkers 2020; 4: 45–59.

28. van Fenema EM, Gal P, van de Griend MV et al. A pilot study evaluating the physiological parameters of performance-induced stress in undergraduate music students. Digit Biomark 2017; 1: 118–125.

29. Psaltos D, Chappie K, Karahanoglu FI et al. Multimodal wearable sensors to measure gait and voice. Digit Biomark 2019; 3: 133–144.

30. Luo H, Lee PA, Clay I et al. Assessment of fatigue using wearable sensors: A pilot study. Digital biomarkers 2020; 4: 59–72.

31. Jacobson NC, Summers B and Wilhelm S. digital biomarkers of social anxiety severity: Digital phenotyping using passive smartphone sensors. J Med Internet Res 2020; 22: e16875.

32. Silva de Lima AL, Hahn T, Evers LJW et al. Feasibility of large-scale deployment of multiple wearable sensors in parkinson’s disease. PLoS ONE 2017; 12: e0189161.

33. Low CA, Dey AK, Ferreira D et al. Estimation of symptom severity during chemotherapy from passively sensed data: Exploratory study. J Med Internet Res 2017; 19: e420.

34. Saner H, Schuetz N, Buluschek P et al. Case report: Ambient sensor signals as digital biomarkers for early signs of heart failure decompensation. Front Cardiovasc Med 2021; 8: 11.

35. Krizhevsky A, Sutskever I and Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM 2017; 60: 84–90.

36. Deng J, Dong W, Socher R et al. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition. IEEE. ISBN 1424439922, pp. 248–255.

37. Szegedy C, Liu W, Jia Y et al. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, pp. 1–9.

38. He K, Zhang X, Ren S et al. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, pp. 770–778.

39. Gori M, Monfardini G and Scarselli F. A new model for learning in graph domains. In Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005., volume 2. IEEE. ISBN 0780390482, pp. 729–734.

40. Kipf TN and Welling M. Semi-supervised classification with graph convolutional networks, 2016. https://arxiv.org/abs/1609.02907.

41. Mikolov T, Chen K, Corrado G et al. Efficient estimation of word representations in vector space, 2013. https://arxiv.org/abs/1301.3781.

42. Hochreiter S and Schmidhuber J. Long short-term memory. Neural Comput 1997; 9: 1735–1780.

43. Chung J, Gulcehre C, Cho K et al. Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014. https://arxiv.org/abs/1412.3555.

44. Bahdanau D, Cho K and Bengio Y. Neural machine translation by jointly learning to align and translate, 2014. https://arxiv.org/abs/1409.0473.
45. Vaswani A, Shazeer N, Parmar N et al. Attention is all you need. In 31st Conference on Neural Information Processing Systems. NIPS, pp. 5998–6008.

46. Hinton GE and Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006; 313: 504–507.

47. Salakhutdinov R and Hinton G. Deep boltzmann machines. In Proceedings of the 12th International Conference on Artificial Intelligence and Statistics. PMLR, pp. 448–455.

48. Rumelhart DE, Hinton GE and Williams RJ. Learning internal representations by error propagation. ICS report 8506. La Jolla, CA: University of California, 1985.

49. Vincent P, Larochelle H, Lajoie I et al. Stacked denoising auto-encoders: Learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res 2010; 11: 3371–3408. <GotoISI>://WOS:000286637200004.

50. Brisk R, Bond R, Finlay D et al. Better-than-expert detection of early coronary artery occlusion from 12 lead electrocardiograms using deep learning, 2019. https://www.researchgate.net/publication/331670484_Better-than-expert_detection_of_early_coronary_artery_occlusion_from_12_lead_electrocardiograms_using_deep_learning.

51. Hwang B, You J, Vaessen T et al. Deep ecgnet: An optimal deep learning framework for monitoring mental stress using ultra short-term ecg signals. Telemedicine and E-Health 2018; 24: 753–772.

52. Attia ZI, Sugrue A, Asrivatham SJ et al. Noninvasive assessment of dofetilide plasma concentration using a deep learning (neural network) analysis of the surface electrocardiogram: A proof of concept study. PLoS ONE 2018; 13: e0201059.

53. Attia ZI, Noseworthy PA, Lopez-Jimenez F et al. An artificial intelligence-enabled ecg algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet 2019; 394: 861–867.

54. Erdenebayar U, Kim H, Park JU et al. Automatic prediction of atrial fibrillation based on convolutional neural network using a short-term normal electrocardiogram signal. J Korean Med Sci 2019; 34: e64.

55. Urtnasan E, Park JU, Joo EY et al. Identification of sleep apnea severity based on deep learning from a short-term normal ecg. J Korean Med Sci 2020; 35: e399.

56. Sridhar N, Shobh A, Stephens P et al. Deep learning for automated sleep staging using instantaneous heart rate. NPJ digital medicine 2020; 3: 1–10.

57. Xiao R, Xu Y, Pelter MM et al. A deep learning approach to examine ischemic st changes in ambulatory ecg recordings. AMIA Jt Summits Transl Sci Proc 2018; 2017: 256–262. https://www.ncbi.nlm.nih.gov/pubmed/29888083.

58. Lui HW and Chow KL. Multiclass classification of myocardial infarction with convolutional and recurrent neural networks for portable ecg devices. Informatics in Medicine Unlocked 2018; 13: 26–33.

59. Seo W, Kim N, Kim S et al. Deep ecg-respiration network (deeper net) for recognizing mental stress. Sensors (Basel) 2019; 19: 3021.

60. Zhang H, Deng K, Li H et al. Deep learning identifies digital biomarkers for self-reported parkinson’s disease. Patterns (N Y) 2020; 1: 100042.

61. Siebets SK, Schaff J, Duda M et al. Crowdsourcing digital health measures to predict parkinson’s disease severity: the parkinson’s disease digital biomarker dream challenge. NPJ Digit Med 2021; 4: 53.

62. McClure K, Erdreich B, Bates JHT et al. Classification and detection of breathing patterns with wearable sensors and deep learning. Sensors (Basel) 2020; 20: 6481.

63. Lee KS, Chae S and Park HS. Optimal time-window derivation for human-activity recognition based on convolutional neural networks of repeated rehabilitation motions. In 2019 IEEE 16th international conference on rehabilitation robotics (ICORR). IEEE. ISBN 1728127556, pp. 583–586.

64. Varghese J, Niewohner S, Soto-Rey I et al. A smart device system to identify new phenotypical characteristics in movement disorders. Front Neurol 2019; 10: 48.

65. Torvi VG, Bhattacharya A and Chakraborty S. Deep domain adaptation to predict freezing of gait in patients with parkinson’s disease. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE. ISBN 153866805X, pp. 1001–1006.

66. Mondal R, Mukherjee D, Singh PK et al. A new framework for smartphone sensor-based human activity recognition using graph neural network. IEEE Sens J 2020; 21: 11461–11468.

67. Kim K, Lee S, Kim S et al. Deep learning-based abnormal behavior detection system for dementia patients. Journal of Internet Computing and Services 2020; 21: 133–144.

68. Ahmed N, Singh A KSS et al. Classification of sleep-wake state in a ballistocardiogram system based on deep learning, 2020. https://arxiv.org/abs/2011.08977.

69. Kim HM, An JI and Park YR. A prediction model for detecting developmental disabilities in preschool-age children through digital biomarker-driven deep learning in serious games: Development study. JMIR Serious Games 2021; 9: e23130.

70. Srivastava A, Jain S, Miranda R et al. Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease. PeerJ Comput Sci 2021; 7: e369.

71. Ahmed MY, Rahman MM and Kuang J. Deepplung: Smartphone convolutional neural network-based inference of lung anomalies for pulmonary patients. In INTERSPEECH. INTERSPEECH, pp. 2335–2339.

72. Bidja P. Depressiongnn: Depression prediction using graph neural network on smartphone and wearable sensors. Honors Scholar Theses 2019.

73. Zhang W, Wang F, Wu S et al. Partial directed coherence based graph convolutional neural networks for driving fatigue detection. Rev Sci Instrum 2020; 91: e074713.

74. Ozbekcan Y, Ozturk MG, Ekmecki MN et al. Depression screening from voice samples of patients affected by parkinson’s disease. Digit Biomark 2019; 3: 72–82.

75. Pyrkov TV, Sokolov IS and Fedichev PO. Deep longitudinal phenotyping of wearable sensor data reveals independent markers of longevity, stress, and resilience. Aging (Albany NY) 2021; 13: 7900–7913.

76. Jh Jung, Dh Lee and al Kim Ss et. Deep learning-based user emergency event detection algorithms fusing vision, audio, activity and dust sensors. J Internet Comput Serv 2020; 21: 109–118.

77. Rutkowski TM, Koculak M, Abe MS et al. Brain correlates of task–load and dementia elucidation with tensor machine
learning using oddball bci paradigm. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE. ISBN 1479981311, pp. 8578–8582.

78. Han SY, Kwak NS, Oh T et al. Classification of pilots’ mental states using a multimodal deep learning network. Biocybernetics Biomed Eng 2020; 40: 324–336.

79. Dong G, Cai L, Datta D et al. Influenza-like symptom recognition using mobile sensing and graph neural networks. In Proceedings of the Conference on Health, Inference, and Learning. CHIL. pp. 291–300.

80. Merrill MA and Althoff T. Transformer-based behavioral representation learning enables transfer learning for mobile sensing in small datasets. 2021. https://arxiv.org/abs/2107.06097.

81. Lee H, Lee W and Jeong T. Deep learning algorithm and prediction model associated with data transmission of user-participating wearable devices. Journal of the Korea Industrial Information Systems Research 2020; 25: 33–45.

82. Kanjo E, Younis EMG and Ang CS. Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. Inf Fusion 2019: 49: 46–56.

83. Mou LT, Zhou C, Zhao PF et al. Driver stress detection via multimodal fusion using attention-based cnn-lstm. Expert Syst Appl 2021; 173: 114693.

84. He L, Jiang D, Yang L et al. Multimodal affective dimension prediction using deep bidirectional long short-term memory recurrent neural networks. In Proceedings of the 5th Workshop on Audio/Visual Emotion Challenge. AVEC, pp. 73–80.

85. Oseledets IV. Tensor-train decomposition. SIAM J Sci Comput 2011; 33: 2295–2317.

86. Hamilton WL, Ying R and Leskovec J. Inductive representation learning on large graphs. In Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS, pp. 1025–1035.

87. Radu V, Lane ND, Bhattacharya S et al. Towards multimodal deep learning for activity recognition on mobile devices. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. ACM, pp. 185–188.

88. Suk HI, Lee SW, Shen D et al. Hierarchical feature representation and multimodal fusion with deep learning for ad/inci diagnosis. Neuroimage 2014; 101: 569–582.

89. Alam MR, Bennamoun M, Togneri R et al. A joint deep boltzmann machine (jdbm) model for person identification using mobile phone data. IEEE Trans Multimedia 2016; 19: 317–326.

90. Liu HY, Deng SU, Wu LF et al. Recommendations for different tasks based on the uniform multimodal joint representation. Applied Sciences-Basel 2020; 10: 6170 10.3390/app10186170. <GotoISI>://WOS:000581774600001.

91. Khamparia A, Saini G, Pandey B et al. Kdsae: Chronic kidney disease classification with multimedia data learning using deep stacked autoencoder network. Multimed Tools Appl 2020; 79: 35425–35440.

92. Said AB, Mohamed A, Elfouly T et al. Multimodal deep learning approach for joint eeg-eng data compression and classification. In 2017 IEEE wireless communications and networking conference (WCNC). IEEE. ISBN 1509041834, pp. 1–6.

93. Du LH, Liu W, Zheng WL et al. Detecting driving fatigue with multimodal deep learning. In 2017 8th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE. ISBN 1509046038, pp. 74–77.

94. Liu YN, Feng XQ and Zhou ZG. Multimodal video classification with stacked contractive autoencoders. Signal Processing 2016; 120: 761–766.

95. Wang S, Zhang J and Zong C. Associative multichannel autoencoder for multimodal word representation. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, pp. 115–124.

96. Zhang P, Ma X, Zhang W et al. Multimodal fusion for sensor data using stacked autoencoders. In 2015 IEEE tenth international conference on intelligent sensors, sensor networks and information processing (ISSNIP). IEEE. ISBN 1479980552, pp. 1–2.

97. Henderson N, Emerson A, Rowe J et al. Improving sensor-based affect detection with multimodal data imputation. In 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE. ISBN 1728138884, pp. 669–675.

98. Liu W, Qiu JL, Zheng WL et al. Multimodal emotion recognition using deep canonical correlation analysis. 2019. https://arxiv.org/abs/1908.05349.

99. Lan YT, Liu W and Lu BL. Multimodal emotion recognition with multimedia data learning and information processing (ISSNIP). IEEE. ISBN 1479980552, pp. 1–2.

100. Yang X, Ramesh P, Chitta R et al. Deep multimodal representation learning from temporal data. In Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, pp. 5447–5455.