AI-Based Stroke Disease Prediction System Using ECG and PPG Bio-Signals

JAEHAK YU[1], SEJIN PARK[2], SOON-HYUN KWON[1], KANG-HEE CHO[3], AND HANSUNG LEE[4]

[1]Department of Knowledge-Converged Super Brain (KSB) Convergence Research, Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea

[2]Safety Measurement Institute, Korea Research Institute of Standards and Science, Daejeon 34113, South Korea

[3]Department of Rehabilitation Medicine, College of Medicine, Chungnam National University, Daejeon 35015, South Korea

[4]School of Creative Convergence, Andong National University, Andong, Gyeongsangbuk-do 36729, South Korea

Corresponding author: Hansung Lee (mohan@anu.ac.kr)

ABSTRACT Since stroke disease often causes death or serious disability, active primary prevention and early detection of prognostic symptoms are very important. Stroke diseases can be divided into ischemic stroke and hemorrhagic stroke, and they should be minimized by emergency treatment such as thrombolytic or coagulant administration by type. First, it is essential to detect in real time the precursor symptoms of stroke, which occur differently for each individual, and to provide professional treatment by a medical institution within the proper treatment window. However, prior studies have focused on developing acute treatment or clinical treatment guidelines after the onset of stroke rather than detecting the prognostic symptoms of stroke. In particular, in recent studies, image analysis such as magnetic resonance imaging (MRI) or computed tomography (CT) has mostly been used to detect and predict prognostic symptoms in stroke patients. Not only are these methodologies difficult to diagnose early in real-time, but they also have limitations in terms of a long test time and a high cost of testing. In this paper, we propose a system that can predict and semantically interpret stroke prognostic symptoms based on machine learning using the multi-modal bio-signals of electrocardiogram (ECG) and photoplethysmography (PPG) measured in real-time for the elderly. To predict stroke disease in real-time while walking, we designed and implemented a stroke disease prediction system with an ensemble structure that combines CNN and LSTM. The proposed system considers the convenience of wearing the bio-signal sensors for the elderly, and the bio-signals were collected at a sampling rate of 1,000Hz per second from the three electrodes of the ECG and the index finger for PPG while walking. According to the experimental results, C4.5 decision tree showed a prediction accuracy of 91.56% while RandomForest showed a prediction accuracy of 97.51% during walking by the elderly. In addition, the CNN-LSTM model using raw data of ECG and PPG showed satisfactory prediction accuracy of 99.15%. As a result, the real-time prediction of the elderly stroke patients simultaneously showed high prediction accuracy and performance.

INDEX TERMS Deep learning, machine learning, electrocardiogram (ECG), photo plethysmography (PPG), multi-modal bio-signal, real-time stroke prediction, stroke disease analysis.

I. INTRODUCTION

Stroke can be categorized into ischemic stroke, in which a blood vessel supplying blood to a part of the brain is blocked,

The associate editor coordinating the review of this manuscript and approving it for publication was Md Kafiul Islam.

and hemorrhagic stroke, in which a blood vessel bursts. It is a neurological symptom and disease caused by damage to the brain in a particular area [1]–[5]. Stroke is considered one of the most serious diseases in modern society as it can cause death in severe cases, while also leading to physical and mental disorders such as hemiparesis, speech impairment

(aphasia), ataxia, visual impairment, consciousness impairment, and dementia [6]. According to the 2019 Causes of Death Report released by the World Health Organization (WHO) in December 2020, the top 10 causes of death accounted for 55% of all recorded deaths in 2019 (about 55.4 million people). Among them, 6 million deaths were due to cerebrovascular disease, which was reported to be the second leading cause of death [7]. The United Nations (UN) reported that a country is classified as an aging society when the proportion of its population aged 65 and over in the total population is 7% or more, an aged society when the proportion is 14% or more, and a super-aged society when the proportion is over 20% [8]. As such, the social problems of the aging society are becoming prominent enough that the aging society can be analyzed through segmentation. In addition, according to an analysis report on aging by Moody's, an international credit rating company, as of 2013, Japan, Germany, Italy, etc. had become super-aged society, wherein the proportion of elderly people is over 20%. It has been reported that by 2030, a whopping 34 countries will have become super-aged societies [8]. The prognosis and health status of patients after stroke vary substantially with age and location of onset. According to a previous study on stroke, more than 66% of total stroke incidence occurred in elderly people 65 years or older [9]. In addition to these social problems, the incidence and mortality of stroke are expected to emerge as important social and economic issues.

The diagnosis of stroke, which is represented by cerebrovascular disease, is judged by a medical team's neurological diagnosis and information on severity [6], [10]–[12]. The main methods used for neurological diagnosis in stroke diagnosis are brain MRI and CT, while other studies have reported that bio-signals such as brain waves, muscle, and electrocardiogram can also be used to diagnose and prevent stroke diseases [13]–[15]. In addition, ultrasound examination, echocardiography, cerebral angiography, and single photon emission computed tomography (SPECT) are being used to determine the common causes of stroke [16]. Recently, imaging techniques such as CT and MRI have been widely used for stroke diagnosis, but these still have inconveniences in the examination and diagnosis process caused by hypersensitive reactions according to the drug penetration of the contrast agent, radiation exposure, and claustrophobia in a confined space. There may also be errors in the test results, so judgment based on the medical staff's professional medical knowledge and empirical evidence is considered to be very important. Next, the national institutes of health stroke scale (NIHSS), which is published by the US national institute of health, is used as a research method to prevent recurrence and evaluate early disabilities in stroke patients [6], [17]–[19]. Although the NIHSS is the methodology used to evaluate initial impairment in stroke patients, there are still limitations with detecting initial impairment in real-time and difficulty in clinical and psychological analyses. Recent research has shown that many studies have used ECG to predict and prevent stroke diseases through atrial fibrillation, one of the key causes of stroke [20], [21]. Atrial fibrillation (AF) is an independent major risk factor for stroke that often accompanies hypertensive patients, and it is known to increase the risk of cerebral infarction by more than 5 times [22]. A prior study used clinical trials to identify risk factors for stroke diseases, and they reported that these factors included high blood pressure, smoking, obesity, and diabetes [23], [24]. Therefore, there is a need for a way for the elderly to evaluate individual stroke risk factors and detect the possibility of disease early in real-time. To overcome these limitations, studies have recently attempted to predict stroke diseases using statistical or machine learning methodologies considering certain risk factors. However, there is a limitation that these methodologies also provide black-box results for predictive results, thus making them difficult to interpret. For decision tree methodologies, partial analytical approaches can be used with heuristic methodologies, but studies on predictive systems that can cover the methodologies of prior studies while also encompassing the possibility of developing disease and semantic interpretations are currently highly desirable.

In this paper, we propose a new system that provides stroke disease prediction and semantic interpretation results for the elderly based on ECG and PPG-based multi-modal bio-signals. The proposed system can instantly detect and predict the prognostic symptoms of stroke disease in the elderly by collecting multi-modal bio-signals in real-time. The participants in this paper were people aged 65 or older, and multi-modal bio-signals of ECG and PPG were collected and stored while these participants were walking. The collected multi-modal bio-signals data are attributes divided by signal waveforms into specific intervals, which are intended to be used to study predictive models in machine learning and make relatively accurate predictive results and semantic interpretations. It was also experimentally verified that deep learning time series analysis models can accurately detect stroke prognostic symptoms by using raw data as is, without using separate attribute extraction and features. The multi-modal bio-signals-based disease prediction system for the elderly proposed in this paper can detect and predict in real-time the prognostic symptoms of stroke with high mortality and incidence rate. In this paper, we defined and discovered 29 new attributes that were not previously used in ECG and PPG multi-modal bio-signals based studies in machine learning and deep learning techniques. This is a major contribution in that it can be actively used for objective diagnosis and prognostic treatment by providing semantic analysis results to medical staff. It was experimentally verified that the stroke prediction and monitoring system described in this paper can be used for the real-time prediction of prognostic symptoms of stroke disease, as well as in low-cost daily life health care services. The data used in this experiment were the bio-signals of ECG and PPG collected in real-time while elderly people aged 65 years or older were walking. Among the 29 meaningful attributes and various machine learning techniques used, RandomForest showed a prediction accuracy of 98.31%. In addition, the deep learning

CNN-LSTM model confirmed a satisfactory prediction accuracy of 99.15%, which means that it simultaneously presented high performance for the stroke diseases prediction system.

The structure of the rest of this paper is as follows. Section 2 examines prior research on machine learning and stroke diseases using multi-modal bio-signals from ECG and PPG. Section 3 describes the real-time multi-modal bio-signals-based stroke disease prediction and monitoring system proposed in this paper. Section 4 details the experimental results obtained, and the semantic analysis and discussion is summarized in Section 5. Section 6 finally discusses the conclusions and directions of future research.

II. RELATED WORKS
A. STROKE DISEASE STUDY USING ECG

Electrocardiogram is a bio-signal measured by electrodes and equipment attached to the skin that can be used to analyze the electrical activity of the heart for a set period of time [25]–[27]. ECG bio-signals are used as an important means of nonvascularized diagnosis of the state of the heart, and they are one of the typical biopotential signals based on the number of amplitudes and frequencies. ECG is used not only to measure the rate and consistency of heartbeats, but also to capture cardiac modulators such as heart size and location, heart damage, pacemaker, and effects such as those of medicine [25]. In addition to x-ray testing, ECG is a vital test for the diagnosis of heart disease, and particularly for cardiac arrhythmias with irregular heartbeats. ECG is also useful for diagnosing cardiomyopathy, atrial or ventricle hypertrophy, dilation, pulmonary circulation disorders, electrolyte metabolic abnormalities, drug effects, and other heart and associated diseases [25]. ECG is a non-vascular test that has the advantage of being relatively easy to repeat without causing pain to patients [25], [26].

According to a recent preliminary study, most studies using ECG are focused on detecting cardiac arrhythmia or AF [28]–[35]. The prior studies that detected and classified AF [28]–[31] or arrhythmia [32]–[35] using ECGs include the following: Couceiro *et al.* [28] conducted a study to extract the characteristics of AF from ECG using three major physiological and clinical attributes of AF. Specifically, they analyzed the three characteristics of P wave absence, heart rate irregularity, and atomic activity, and they reported that their sensitivity and specificity were 93.80% and 96.09%, respectively. These studies confirm that AF symptoms can be detected with relatively high accuracy. Xiong *et al.* [29] designed and studied deep learning-based 1D CNNs consisting of 16 layers to classify ECGs including AF. These techniques are methods of skip interlayer connection between layers. It has been reported that the training time for learning and the speed of transmitting the weight values of the connection layer can be improved as well. Experiments have also shown that normal rhythms and other rhythms, including AF, could be accurately classified within 0.01 seconds. For example, Bumgarner *et al.* [30] studied Kardia Band (KB), a new

technique by which rhythm strips could be recorded using an Apple Watch (Apple, Cupertino, California). Through an app that provides automatic detection of atrial fibrillation, an experiment was conducted to determine whether AF and sinus rhythm can be accurately distinguished compared to 12-lead ECG and KB records interpreted by medical staff. Of the 57 non-interpretable KB data, the interpretation of electrophysiologists confirmed that they could diagnose AF with a sensitivity of 100%, a specificity of 80%, and a K coefficient of 0.74.

These experimental results indicate that medical staff can accurately distinguish and detect AF and sinus rhythm using KB. The study by Bumgarner *et al.* also indicate that the accurate diagnosis of patients before cardioversion is selected may be helpful.

B. DISEASE PREDICTION STUDY USING PPG

The cardiovascular system consists of the blood vessels responsible for blood circulation in the body and the heart. Cardiovascular diseases include arrhythmias, myocardial infarction, and heart failure, while vascular diseases include alliance sclerosis and hyperlipidemia [36]. Although imaging techniques such as ultrasound, MRI, and digital subtraction angiography are used as diagnostic methods for cardiovascular disease, they have certain disadvantages such as a long measurement time, high cost, and radiation exposure. PPG is a bio-signal that compensates for these shortcomings and considers the limitations of conducting measurements during daily activities [37]–[40]. PPG is measured using the linear relationship between the blood volume, which changes with the contraction and relaxation of the heart, as well as the light absorbed by hemoglobin. This is a simple, non-invasive, low-cost method with a fast measurement time that can be used for the early prediction of cardiovascular disease [37]–[39].

Nichols *et al.* [37] experimentally showed that aortic pulse wave velocity and augmentation index were independent predictors of adverse cardiovascular events including mortality. They also verified experimentally that in hypertension and aging, the central elastic arteries stiffened and the diastolic pressure decreased, leading to increased central systolic and pulse pressures. As a result of these symptoms, left ventricular (LV) energy remarkably increased, and pulse wave velocity (PWV) and pulse pressure amplification increased as well. Therefore, the LV load and the myocardial oxygen demand are increased, causing a mismatch of the arterial pulse wave, and thereby promoting ventricular hypertrophy. As a result, the inconsistency of the arterial pulse wave and the onset of atherosclerosis are symptoms that lead to cerebral infarction and stroke as well as cardiovascular diseases, so rapid and accurate treatment and diagnosis are essential research topics. Allen *et al.* [38] defined PPG as a simple and low-cost optical technique that can be used to detect blood volume changes in the microvascular bed of tissue. The measurement protocol and pulse wave analysis use data collected from the fingers, ears, and toes, as these are locations where PPG bio-signals can be experimentally measured. The PPG bio-signals results

indicate that it can be used in healthcare services by measuring clinical physiology, including clinical physiological monitoring, vascular assessment, and autonomic function. Allen *et al.* introduced comparison and analysis information according to ECG and PPG measurement location; the prospects for analytical techniques including PPG images and simple endothelial dysfunction assessment and home diagnosis through experiments and validation were detailed as well. Oliver *et al.* [39] reported that pulse wave velocity of PPG, diameter, or area of an artery can be used as a noninvasive assessment method in arterial stiffness and large arterial cases. As a result, abnormalities in cardiovascular diseases such as arteriosclerosis using PPG were checked, and the effects of future cardiovascular drugs on arteriosclerosis were analyzed and reported. Altogether, PPG has been widely used to detect changes in blood pressure and discrepancies in arterial pulses by analyzing pulse transit time, which is a variation in the vibration width of arterial vessels. However, because PPG alone detects and estimates the correlation between vascular stiffness, arterial blood pressure, and changes in blood pressure, there is a limit in predicting the onset of atherosclerosis and identifying symptoms that lead to stroke. Therefore, there is a need for a method capable of quickly and accurately detecting prognostic symptoms leading to cerebral infarction and stroke as well as cardiovascular diseases by simultaneously examining the waveform of ECG and PPG rather than PPG alone.

C. STROKE RESEARCH BASED ON ARTIFICIAL INTELLIGENCE TECHIQUE

According to a literature review, a number of studies have reported the detection of abnormalities such as cardiovascular diseases and arrhythmias using AI techniques and various bio-signals [32], [33]. Recently, studies have been actively conducted to predict the condition of heart disease based on ECG or pulse among bio-signals collected in real-time. For example, Sannino *et al.* [32] performed an experiment using the MIT-BIH arrhythmia database [41] for ECG classification and analyzed the results. They proposed an approach based on a deep neural network for automatic classification of normal and abnormal ECGs. The DNN model was composed of 7 hidden layers, and the neurons for each layer were set to 5, 10, 30, 50, 30, 10, and 5. For comparison with existing machine learning algorithms, 11 classification algorithms were selected using WEKA [42], an open package of data mining, and a comparison experiment was performed. The experimental results showed that the learning accuracy of the proposed model was 100%, the test accuracy was 99.09%, and the accuracy of the entire data was 99.68%, which verified that normal and abnormal heartbeats can be classified with very high performance. Yıldırım *et al.* [35] designed a deep learning-based 1D-CNN model that can efficiently detect arrhythmia in heart disease in real-time. In this experiment, one lead was used, and the classification accuracy of 91.33% was reported for each of 17 classes including normal

sinus rhythm, cardiac arrhythmia, and pacemaker rhythm for 10 seconds of ECG.

Recently, a number of studies have attempted to detect abnormalities in heartbeats using ECG and predict stroke diseases based on such abnormalities [43]–[47]. Kamel *et al.* [43] conducted a multi-ethnic study of 6,741 atherosclerosis participants without cerebrovascular disease or AF. For ECG measurement, the associations between left atrial aberrations markers were measured and experimented with standard 12-lead methods. During the median follow-up period of 8.5 years, it was reported that 121 participants (1.8%) were diagnosed with stroke, while 541 patients (8.0%) were diagnosed with AF. Based on analysis during the follow-up period, the association between P-wave morphology and stroke disease was newly discovered. He *et al.* [44] estimated the overall effect by applying a random-effects models to data up to July 2017 for a study evaluating the association between P-wave morphology and stroke disease. P-wave duration in V1 lead was shown to be a significant predictor of stroke disease when analyzed as a categorical variable, but it was not predictable when analyzed as a continuous variable. It was experimentally verified that the maximum P-wave region can also predict the risk of ischemic stroke. According to the experimental results, the P-wave duration and maximum P-wave region in V1 lead can be used to predict the risk of ischemic stroke disease.

III. REAL-TIME STROKE DISEASE PREDICTION SYSTME BASED ON MULTIPLE BIO-SIGNALS FROM ECG AND PPG

In this paper, we propose a system for monitoring stroke disease and health among the elderly based on multimodal bio-signals of ECG and PPG collected in real-time. The proposed system extracts attributes based on the peak values of waveforms from raw data of ECG and PPG, then applies a machine learning algorithm to predict the prognostic symptoms of stroke in the elderly in real-time. In addition, the raw data of ECG and PPG are applied to the deep learning model to accurately predict the prognostic symptoms of stroke disease in real-time. The structure of the proposed system includes: 1) a bio-signal sensor measurement and transmission module that collects bio-signals such as ECG and PPG in real-time; 2) a module that collects, stores, and transmits multimodal bio-signals generated in real-time to a server; 3) a module for extracting and updating important attributes based on stored bio-signals such as ECG and PPG; 4) a module for learning based on a machine learning algorithm using property information for each bio-signals as well as a deep learning model using raw data; and 5) a visualization module that provides the stroke prognostic symptoms and prediction results of the elderly to medical staff or spouses as shown in Figure 1. In other words, the proposed system stores and manages meaningful attributes through measurement, collection, and preprocessing functions for each multimodal bio-signal among ECG and PPG from the elderly and general elderly suffering from stroke. It was also designed to predict and analyze stroke prognostic
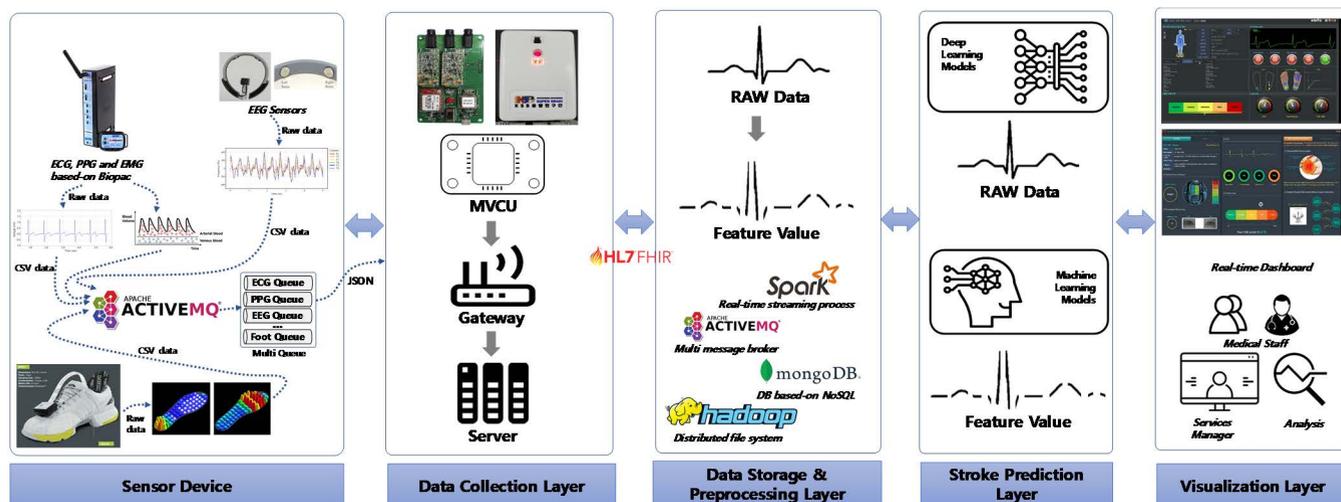
FIGURE 1. Overall architecture of the proposed system.

symptoms in real-time by applying machine learning and preprocessing raw data-based deep learning models using the attributes of each bio-signal.

A. REAL-TIME ECG AND PPG BIO-SIGNALS COLLECTION

In this study, various biological signals have been measured and collected to verify the performance of systems that provide AI-based prognostic and predictive information for older stroke diseases. The collected bio-signals data includes ECG, EEG, PPG, EMG, and motion. This section describes in detail the real-time measurement and collection process. In this study, among the bio-signals measured and collected, EEG and PPG were used in experiments for a performance verification of a system that can predict stroke disease. It has been reported through previous studies that abnormalities in the autonomic nerves and sympathetic nerves appear during the precursor symptoms of stroke or the onset of a stroke. There are many difficulties in accurately predicting stroke symptoms using only a single bio-signal. In this study, important characteristic values were extracted from two types of multimodal bio-signals: ECG, which can confirm the rate and consistency of heartbeat, PPG, which checks the blood volume that changes with contraction and relaxation of the heart. We propose to use a feature that combines two bio-signals to accurately predict stroke prognostic symptoms and onset. The ECG measurement method is largely divided into standard 12-lead ECG and chest guidance depending on the location of attachment, and standard 12-lead ECG can be further divided into bipolar standard guidance and unipolar extremity guidance. In this study, the ECGs of elderly stroke patients and general elderly were precisely measured and collected using the chest guidance method. The three electrode attachment positions of the ECG used in the experiments in this paper are illustrated in detail in Figure 2. The PPG bio-signals measured and collected in this paper were stored in real-time
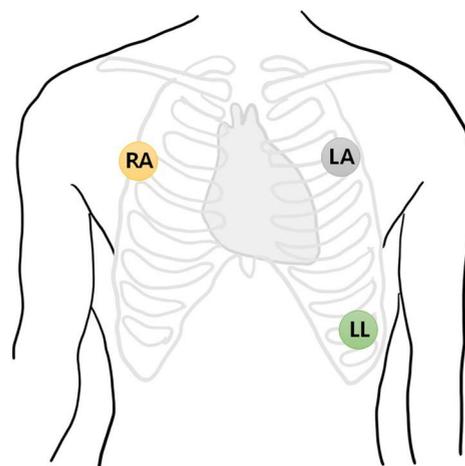
FIGURE 2. 3-electrode-based measurement and acquisition for ECG bio-signal acquisition.

by fixing the sensors to the left and right index fingers of the subject, as shown in Figure 3.

These EEG and PPG bio-signals were collected from subjects aged 65 years or older from 2017 to 2018 at the emergency medical center and department of rehabilitation medicine at Chungnam National University Hospital, Republic of Korea. The subjects were patients who had been diagnosed with a stroke within one month at Chungnam National University Hospital and who were receiving treatment at the Department of Neurology and Rehabilitation Medicine, Chungnam National University. Subject candidates consisting of stroke patients and normal elderly people were primarily selected by the medical staff. Among the candidates, the subjects determined to be stroke were secondarily verified by the neurologist faculty. In addition to ECG and PPG bio-signals from the subject, the experimental protocol

FIGURE 3. PPG bio-signal measurement and collection.

FIGURE 4. AI-based stroke disease prediction module using multimodal bio-signals.

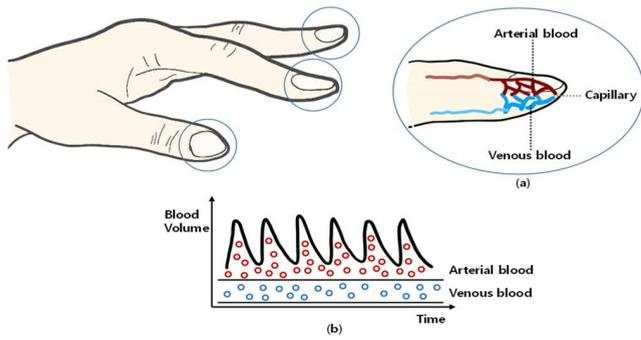was designed to simultaneously collect various bio-signals data through sensors such as EEG, EMG, motion, and foot pressure. In this study, the equipment "Biopac" was used to measure and collect various bio-signals. The synchronization of ECG and PPG raw data was performed by medical staff and they directly monitored and verified the data. The attached bio-signals sensors were checked for normal operation the first time, after which various bio-signals were collected 5 times. Finally, for ECG and PPG bio-signals, 1,000 raw data were measured and collected each second at a sampling rate of 1,000 Hz.

B. AI-BASED STROKE PREDICTION MODULE USING MULTIMODAL BIO-SIGNALS OF ECG AND PPG

The system for providing information about stroke prognostic and prediction information for the elderly using multimodal bio-signals of ECG and PPG measured in real-time consisted of an offline module and an online module. First, offline sub-modules extract preprocessing and important features on raw data from ECGs and PPGs were collected and stored from all subjects. Using feature values through preprocessing, various machine learning and deep learning algorithms were trained to generate prediction models for stroke diseases and provide semantic interpretation information. The online module then provided prognostic and predictive information on stroke in the elderly with multimodal bio-signals and feature values of ECG and PPGs collected in real-time, as shown in Figure 4.

The offline module in Figure 4 consists of a total of five sub-blocks. The first block at the bottom left stores various bio-signals that can be measured in the elderly during daily activities. In addition to ECG and PPG, this block measures, collects, and stores bio-signals such as EEG, EMG, and motion. In the second block, the ECG collected based on three electrodes and the PPG signal, which is collected using the linear relationship between the blood volume changing with the contraction and relaxation of the heart and the amount of light absorbed by hemoglobin in the blood, are pre-processed. The temporarily unmeasured null values are deleted from the raw data values of ECG and PPG, and the remaining data are normalized using the Z-score method. The third sub-block extracts and stores features for machine learning from
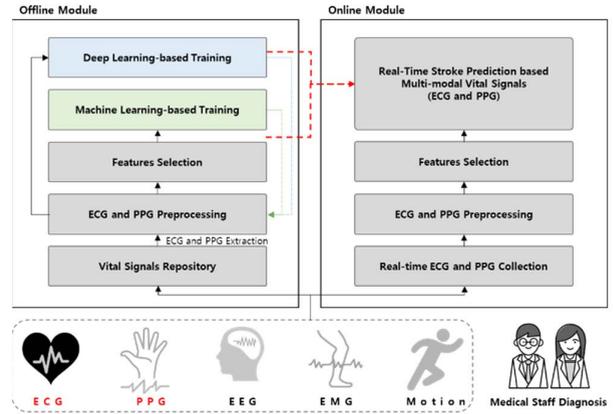
preprocessed raw data, and it creates and stores a trained predictive model. In the fourth block, prediction and analysis of stroke patients and normal elderly people are performed using machine learning methods based on features extracted by ECG and PPG bio-signals. That is, disease prediction and analysis through machine learning methods are sequentially used and processed from the first block to the fourth block. In the fifth block, the deep learning-based stroke prediction model is trained with the preprocessed raw data of ECG and PPG, and this provides the real-time prediction of stroke. In this block, using raw data for each ECG and PPG bio-signal, a deep learning-based prediction model is created, and the model is saved in the form of a meta file.

The online module consists of a total of four sub-blocks. The first sub-block collects real-time ECG and PPG bio-signal data from walking in the daily life of the elderly and stores it. The second sub-block deletes null values from the collected raw data of ECG and PPG, then stores and manages the values applied with the normalization process using Z-score normalization. In the third sub-block, important sections are separated from the raw data of ECG and PPG, and feature values for each bio-signal are extracted from the separated section information. In the fourth block, real-time stroke disease prediction is executed using machine learning and deep learning models with the feature values and raw data extracted from the previous block. The online module can obtain stroke disease prediction results and analysis information of the elderly by using the multimodal bio-signals of EEG and PPG collected in real-time. As a result, the proposed stroke disease prediction results and analysis information can be used as objective data for medical diagnosis and subsequent treatment by medical staff.

C. DEFINITION AND EXTRACTION OF IMPORTANT FEATURE VALUES IN ECG AND PPG BIO-SIGNALS

In this study, among all the bio-signals collected in various scenarios such as resting state, sleeping, sitting, moving objects, and speaking, only gait data was used. That is, only

TABLE 1. Features and description extracted from ECG and PPG bio-signals.

No.	Feature	Explain
1	ECG_a_P	unit is milliseconds (ms), ms value from a to P of ECG
2	ECG_P_b	unit is ms, ms value from P to b of ECG
3	ECG_b_c	unit is ms, ms value from b to c in ECG
4	ECG_c_Q	unit is ms, ms value from c to Q of ECG
5	ECG_Q_R	unit is ms, ms value from Q to R of ECG
6	ECG_R_S	unit is ms, ms value from R to S in ECG
7	ECG_S_d	unit is ms, ms value from S to d in ECG
8	ECG_d_e	unit is ms, ms value from d to e in ECG
9	ECG_PR_width	unit is ms, ms, or width values from P to R in ECG
10	ECG_e_T	unit is ms, ms value from e to T in ECG
11	ECG_S_T	unit is ms, ms value from S to T in ECG
12	ECG_P_peak	unit is ± mv, P peak value of ECG
13	ECG_Q_peak	unit is ± mv, Q peak value of ECG
14	ECG_R_peak	unit is ± mv, R peak value of ECG
15	ECG_S_peak	unit is ± mv, S peak value of ECG
16	ECG_T_peak	unit is ± mv, E peak value of ECG
17	ECG_RRI	unit is ms, ms value from R to next R
18	PPG_L_a_b	ms values from a to b of the left hand PPG
19	PPG_L_a_c	ms values from a to c of the left hand PPG
20	PPG_L_b_d	ms values from b to d of the left hand PPG
21	PPG_L_b_c	ms values from b to c of the left hand PPG
22	PPG_L_b_peak	peak value of b in left hand PPG
23	PPG_R_a_b	ms values from a to b of the right hand PPG
24	PPG_R_a_c	ms values from a to c of the right hand PPG
25	PPG_R_b_d	ms values from b to d of the right hand PPG
26	PPG_R_b_c	ms values from b to c of the right hand PPG
27	PPG_R_b_peak	peak value of b in right hand PPG
28	ECGd_PPGb_L	ms values from left hand to d in ECG and b in PPG
29	ECGd_PPGb_R	ms values from right hand to d in ECG and b in PPG
30	Class Labeling	normal elderly or stroke elderly

FIGURE 5. ECG signal waveform and information on the location of each attribute in the waveform.

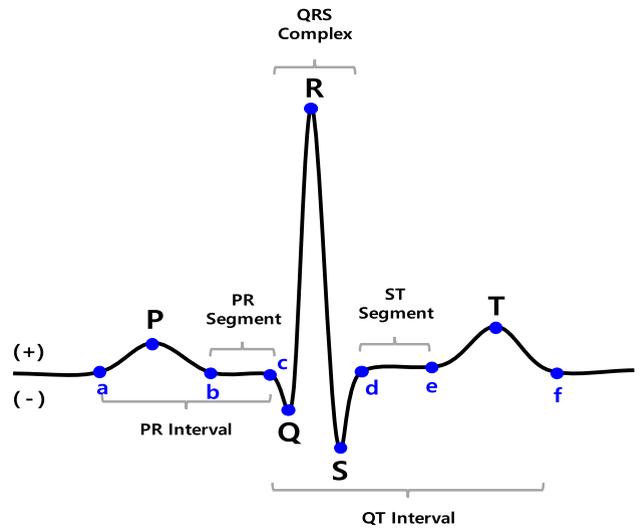**FIGURE 6.** PPG signal waveform and each attribute in the waveform.

the bio-signals of ECG and PPG limited to walking scenarios were used to predict and analyze stroke disease in the elderly. Table 1 describes in detail the 29 attributes and their meanings extracted from the ECG data illustrated in Fig. 5 and the PPG bio-signals shown in Fig. 6.

In this study, Z-score normalization of Equation (1) is applied to all 29 attributes. Since the categories and sizes of the minimum and maximum values of the ECG_a_P attribute vary between patients, there is a problem dependent on the measurement unit, so it is necessary to prevent this. This normalization process transforms the data so that it is located within a small range such as 0.0 to 1.0, so that the same weight is applied to all attributes.
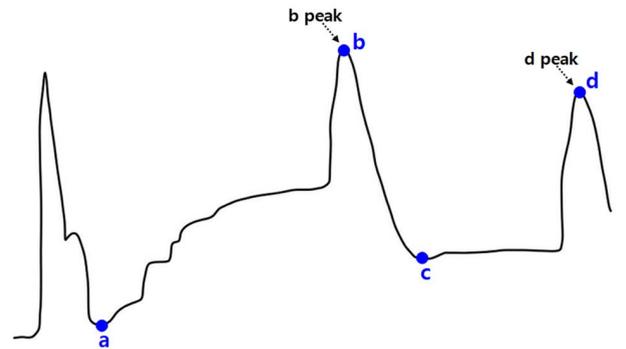
$$\vec{x_i} = \frac{x_i - \mu}{\sigma} \times \alpha \qquad (1)$$

where σ and μ are the standard deviation nd mean of attribute x, respectively. The weight value α is set to 1.0.

When all attributes are used for pattern classification, the execution time of the pattern recognizer is long and the classifier performance is sometimes poor [48], [49]. Therefore, it is essential to efficiently select attribute subsets to reduce the dimension of data, shorten execution time, and improve classification performance. Optimal subset selection aims to maintain or increase classifier accuracy while maintaining

minimal performance degradation so that it can be used more quickly and efficiently. This paper uses the method described by Hall [49]. This methodology calculates the conditional probability using best-first search, entropy for attribute values, and Pearson's correlation coefficient between target class attributes, then finds the minimum number of feature sets in which the probability distribution of all attributes can be expressed adjacently. To obtain the information benefit of each attribute, the entropy for an arbitrary attribute Y is calculated as in Equation (2) [48], [49].

$$H(Y) = -\sum_{y \in Y} p(y) \, log_2(p(y)) \qquad (2)$$

A merit function (Equation (3)) is used to evaluate how efficiently each subset $F_s \subset F$ represents all attributes. Here, it indicates that the subset with the largest value of the merit function is the subset that can optimally express all attributes [49].

$$Merit(F_s) = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \qquad (3)$$

where k denotes the number of attributes in the subset F_s, $\overline{r_{cf}}$ denotes the average distribution of attributes included in F_s, and $\overline{r_{ff}}$ denotes the average correlation value of attributes. In this paper, through the merit function method proposed by Hall [49], 12 optimal attribute subsets among 29 features were selected. The selected 12 properties are {ECG_a_P, ECG_Q_R, ECG_R_S, ECG_PR_width, ECG_S_T, ECG_R_peak, ECG_S_peak, ECG_RRI, PPG_L_b_peak, PPG_R_a_b, PPG_R_b_d, and PPG_R_b_peak}.

IV. EXPERIMENTAL RESULTS

A. DATA CONFIGURATION AND EXPERIMENTAL DESIGN

This section describes the process of collecting and pre-processing multimodal bio-signals of ECGs and PPGs used for machine learning and deep learning-based stroke disease prediction and in-depth analysis and verification. The bio-signals used in the experiments conducted in this study are multimodal data of ECG and PPG measured and collected in real-time. The bio-signals of ECG and PPG refer to values that can express the contraction and relaxation of the heart, as well as the movement of the electric current passing through the three intestines when the heart beats, which appears in the form of a wavy line. ECG and PPG capture changes in blood vessel volume according to the heartbeat measured in the heart's electrical pathway, heart rhythm, and peripheral regions. By analyzing the bio-signals of ECG and PPG, the occurrence of major cardiovascular diseases such as arrhythmias or atrial fibrillation and chronic diseases such as hypertension can be observed. Therefore, in this study, we utilized the clinical results showing that abnormalities in the autonomic nervous system and sympathetic nervous system can occur due to the prognostic symptoms of stroke. We attempted to predict the prognostic symptoms of stroke disease and interpret the meanings based on the medical opinion and diagnostic results of abnormal symptoms, including cardiac arrest or arrhythmia in ECG and PPG. Further, using the raw data of the bio-signals of ECG and PPG, a deep learning model was used to verify the stroke prognostic and prediction experiments in the elderly.

Multimodal bio-signals of ECG and PPG were measured and collected from the elderly from 2017 to 2018 in the Department of Neurology and Rehabilitation Medicine, and Emergency Medical Center, Chungnam National University, Republic of Korea. The participants were elderly people aged 65 years or older who had been diagnosed with a stroke within the last 1 month, and bio-signals such as ECG and PPG, EEG, EMG, motion, and foot pressure were collected from these participants. Experimental dataset consists of bio-signals collected from 287 stroke elderly patients and 287 normal elderly patients. Data were collected from each 287 elderly patients who had been diagnosed with stroke and were being treated in the rehabilitation department, and who were rehabilitating for other diseases and sources of discomfort other than stroke. In this experiment, patients

TABLE 2. The demographics.

	Mean	Minimum	Maximum	Standard Deviation
Male	74.44	65.0	90.0	6.667
Female	77.82	65.5	99.0	6.661

undergoing rehabilitation for other diseases and discomforts other than stroke were defined as normal elderly. To ensure the objectivity of measuring and collecting bio-signals of all subjects, after wearing bio-signals sensors, measurement was performed a total of five times for each scenario, such as sleeping, standing in a stable state, walking, talking, lifting arms and legs, and sitting and standing. Subjects performed one pre-rehearsal for each scenario. Despite having prior practice, the first measured and collected bio-signals values were not used in the experiment because there is a possibility that human noise may occur due to discomfort in the subject caused by wearing the sensor and the tension state. The last measurement protocol was not reflected in the experimental and performance verification data because it is highly likely that the subject's fatigue will be reflected in the bio-signals data due to their old age and the repetition of experiments. Table 2 shows patients' demographics including their age and gender. The analysis of arrhythmia and symptoms including abnormalities in ECG bio-signals is summarized in Table 3. Table 4 describes the statistical analysis of blood pressure and abnormal values of blood tests measured in the emergency room for subjects who have been diagnosed with stroke.

The data used in this paper consists of two types: 1) The value extracted from the raw data by separating the ECG and PPG bio-signal waveforms of stroke patients and the general elderly by section; 2) Raw data values of ECG and PPG bio-signals, which are directly used for the deep learning approach. For the raw values of the collected ECG and PPG bio-signals, blank values that were not temporarily recorded were removed, and normalization was performed by applying the Z-score technique. ECG and PPG bio-signals for training the model and validation were used in units of 5 sec frame windows; these units were selected because 5 sec frame windows are considered medically and clinically to be the minimum unit necessary for medical personnel to determine conditions such as heart disease based on the waveforms of ECG and PPG. ECG was measured in real-time, and the raw data values for the collected three-electrode-based ECG data were extracted and stored. In the case of PPG data, the raw data values were extracted and collected by placing the sensors on the left and right index fingers. In the machine learning-based experiment, the location and section of the attribute were extracted and used from the raw data of ECG and PPG. Figure 5 shows the ECG bio-signals waveform and the location information of the attributes defined in this paper. The important features of the location and interval values of the attributes of the ECG waveform (refer to Table 1) are then extracted. Since the participants in this study include more

TABLE 3. Statistics and lesion information of stroke subjects.

Characteristics	Number
Gender	
Male	149
Female	138
Age	
65 ~ 69	63
70 ~ 79	120
80 ~ 89	84
≥ 90	8
ECG	
Normal ECG	108
Abnormal ECG	154
Borderline ECG	18
Causes	
Infarction	267
Hemorrhage	16
Transient ischemic attacks	4
Lesions	
Infarction	
Anterior Cerebral Artery	7
Middle Cerebral Artery	164
Posterior Cerebral Artery	25
Basilar artery, Vertebral artery	76
Hemorrhage	
Cortex	4
Basal ganglion	2
Thalamus	4
Brain stem	2
Cerebellum	1
Others	3
Symptom	
Weakness	182
Dysarthria	113
Aphasia	26
Decreased consciousness	40
Facial palsy	21
Headache	12
Dizziness	42
Paresthesia	12

TABLE 4. Medical measurement in the emergency room of the stroke subjects.

Blood pressure test (N=286)	Below range	Normal range	Above range
Diastolic blood pressure (mmHg)	8	87	100
Systolic blood pressure (mmHg)	0	178	199

Vital Sign (N=286)	Normal range	Mean±SD
Diastolic blood pressure (mmHg)	< 80mmHg	84±12.5
Systolic blood pressure (mmHg)	< 120mmHg	154±17.5
Pulse (beat/min)	80 ~ 100	79±14.4
Respiration rate (#/min)	12 ~ 20	20±2
Body temperature (°C)	36.1°C ~ 37.2°C	37±0

Emergency blood test (N=284)	Below range	Normal range	Above range
WBC (103/uL)	3	233	48
RBC (106/uL)	125	158	1
Hb (g/dL)	57	218	9
Hct (%)	66	182	36
Platelet (103/uL)	31	249	4
MCV (fL)	5	264	15
MCH (pg)	3	281	0
MCHC (g/dL)	1	280	3
MPV (fl)	1	262	21
Seg.Neutro (%)	1	211	72
Lymphocyte (%)	80	200	4
Monocyte (%)	19	246	19
Eosinophil (%)	0	282	2
Basophil (%)	0	284	0

normal elderly patients than stroke elderly patients, if all ECG bio-signals of normal elderly patients are used, then there is a high possibility that the trained model and prediction will be biased toward normal elderly patients. Therefore, in this paper, all the 287 ECG bio-signals data of the stroke elderly patients were used, and the same number of data from the normal elderly patients were randomly selected and used. Figure 6 shows the raw data of PPG bio-signals and the location information of the attributes newly defined in this paper. The attributes of ECG and PPG shown in Figures 5 and 6 are used to validate predict models for stroke disease and semantic interpretation by utilizing various machine learning algorithms.

In the machine learning-based experiment, the prediction of stroke symptoms and semantic analysis were performed using 29 attributes extracted from the raw data of multi-modal bio-signals of ECG and PPG (refer to Table 1). The experiments include semantic analysis of stroke disease with machine learning algorithms such as C4.5 Decision Tree, Naïve Bayes, logistic Regression, Multi-Layer Perceptron, Random Forest, Best-First Decision Tree, BayesNet,

AdaBoost, and Support Vector Machine (SVM). In addition, by applying the merit function of Equation (3) described in Section 3.C, we performed prediction and semantic analysis using 12 important features out of 29 features. At this time, experiments with deep learning-based were conducted using the raw data of multi-modal bio-signals of ECG and PPG. We conducted experiments on predicting elderly stroke diseases using the raw data of ECG and PPG that were only preprocessed, and we used deep learning algorithms such as LSTM, Bidirectional LSTM, CNN-LSTM, and CNN-Bidirectional LSTM suitable for time series analysis.

B. PERFORMANCE EVALUATION INDEXES
In this section, the accuracy of predicting stroke prognostic symptoms using the bio-signals of ECG and PPG was verified and the performance of the system was evaluated. In this paper, four performance evaluation indexes were used, and detailed descriptions of each of the indices are provided in Equations (4) through (7) [14], [50], [51]. Accuracy refers to the percentage of the total that correctly predicted elderly with stroke as stroke and normal elderly as normal. F1-score means the harmonic mean of recall and precision, and recall

TABLE 5. Accuracy and F1-score (%) according to machine learning model using all 29 attributes of ECG and PPG.

Datasets / Methods	Train(70%) /Test(30%) Acc.[2]	F1[3]	Train(80%) /Test(20%) Acc.	F1	5-Fold CV[1] Acc.	F1	10-Fold CV Acc.	F1	20-Fold CV Acc.	F1
C4.5 DT[4]	87.42	84.40	88.11	88.10	92.06	92.10	91.07	91.10	92.75	92.80
Naive Bayes	70.52	70.06	71.28	71.50	69.34	69.31	69.15	69.11	70.04	70.01
LR[5]	71.52	71.61	72.27	72.40	70.73	70.70	71.42	71.41	71.13	71.11
MLP[6]	85.43	85.42	89.10	89.03	88.69	88.60	90.77	90.70	89.10	90.71
RF[7]	93.04	93.10	95.54	95.61	96.92	96.90	97.22	97.21	98.31	98.30
BFTree	88.41	88.42	85.64	85.70	89.38	89.40	89.28	89.30	90.37	90.40
BayesNet	74.50	74.50	77.72	77.91	75.09	75.11	76.48	76.51	76.98	77.01
AdaBoost	89.31	89.30	90.31	90.30	90.72	90.70	92.11	92.10	93.73	93.70
SVM	92.62	92.60	93.45	93.40	93.30	93.30	94.70	94.70	94.13	94.12

[1] CV: Cross-Validation, [2] Acc.: Accuracy, [3] F1: F1-score, [4] DT: Decision Tree, [5] LR: Logistic Regression, [6] MLP: Multi-Layer Perceptron, [7] RF: Random Forest.

TABLE 6. Recall and precision (%) according to machine learning model using all 29 attributes of ECG and PPG.

Datasets / Methods	Train(70%) /Test(30%) Recall	Prec.	Train(80%) /Test(20%) Recall	Prec.	5-Fold CV Recall	Prec.	10-Fold CV Recall	Prec.	20-Fold CV Recall	Prec.
C4.5 DT	87.40	87.40	88.10	88.10	92.10	92.10	91.10	91.40	92.80	92.91
Naive Bayes	70.50	70.80	71.30	72.20	69.30	69.40	69.10	69.20	70.02	70.01
LR	71.52	71.63	72.30	73.06	70.70	70.70	71.42	71.41	71.10	71.10
MLP	85.42	85.50	89.10	89.20	88.70	89.50	90.82	91.22	90.80	91.21
RF	93.00	93.21	95.50	95.62	96.90	96.90	97.20	97.20	98.30	98.30
BFTree	88.41	88.40	85.60	85.90	89.60	89.40	89.30	89.40	90.40	90.40
BayesNet	74.50	74.50	77.70	78.51	75.10	75.12	76.52	76.50	77.00	77.01
AdaBoost	89.33	89.32	90.30	90.30	90.70	90.70	92.10	92.10	93.74	93.72
SVM	92.62	92.60	93.40	90.42	93.30	93.30	94.70	94.70	94.12	94.14

is the proportion of stroke patients who have been tested positive. Finally, precision refers to the proportion of subjects who have been tested positive who are actually stroke patients.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (4)$$

where true positive (TP) is the number of data wherein stroke elderly were accurately predicted as stroke elderly, false positive (FP) is the number of data wherein stroke elderly were misclassified as normal elderly, false negative is the number of data wherein normal elderly were misclassified as stroke elderly, and true negative (TN) is the number of data wherein normal elderly were correctly predicted as normal elderly.

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

In the case of disease classification and prediction in the healthcare and medical field, misclassification of a patient with a disease as a general patient without that disease can have a significant impact on aftereffects, and it may put the

life of that patient at risk. Conversely, if a general patient without a given disease is misclassified as having that disease, it may lead to the waste of additional examination costs and treatment time. Therefore, in this paper, the prediction performance of stroke diseases in elderly is comprehensively validated using all four performance evaluation indexes. As a result, the experiment was conducted with a focus on finding and validating a model with high accuracy for predicting stroke prognosis while also having a low false positive rate.

C. EXPERIMENTS AND ANALYSIS WITH MACHINE LEARNING METHODS

An experiment was conducted using 29 attributes separated by a certain section from the ECG and PPG collected in real-time for various machine learning methodologies. In this study, experiments were performed using various machine learning algorithms, including C4.5 Decision Tree, Naïve Bayes, logistic regression, Multi-Layer Perceptron, Random Forest, Best-First Decision Tree, BayesNet, AdaBoost, and SVM. In addition, 12 attributes were selected by applying Equation (3), and the prediction performance was measured using only the selected attributes. In the first experiment, all 29 attributes of ECG and PPG, which separated the

TABLE 7. Accuracy and F1-score (%) according to machine learning model using 12 attributes of ECG and PPG.

Datasets / Methods	Train(70%) /Test(30%)		Train(80%) /Test(20%)		5-Fold CV		10-Fold CV		20-Fold CV	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
C4.5 DT	89.07	89.10	91.08	91.10	90.97	91.00	91.56	91.60	92.75	92.70
Naive Bayes	72.18	72.20	74.25	74.40	71.52	71.50	71.92	71.90	72.51	72.50
LR	72.18	17.22	74.25	74.40	71.13	71.10	71.23	71.20	71.52	71.50
MLP	90.39	90.40	89.60	89.60	88.39	88.40	91.46	91.40	92.06	92.00
RF	95.28	95.30	97.02	92.00	97.21	97.20	97.51	97.70	97.71	97.70
BFTree	89.40	89.40	87.62	87.71	90.17	90.20	90.07	90.10	91.17	91.20
BayesNet	77.15	77.20	76.58	76.60	76.78	76.80	77.08	77.10	77.28	77.30
AdaBoost	90.78	90.79	90.82	90.80	91.83	91.85	92.54	92.60	92.87	92.90
SVM	92.90	92.90	93.78	93.80	94.68	94.70	95.47	95.50	95.62	95.60

TABLE 8. Recall and precision (%) according to machine learning model using 12 attributes of ECG and PPG.

Datasets / Methods	Train(70%) /Test(30%)		Train(80%) /Test(20%)		5-Fold CV		10-Fold CV		20-Fold CV	
	Recall	Prec.	Recall	Prec.	Recall	Prec.	Recall	Prec.	Recall	Prec.
C4.5 DT	89.10	89.10	91.10	91.10	91.00	91.10	91.60	91.90	92.80	93.00
Naive Bayes	72.20	72.30	74.32	75.31	71.50	71.50	71.90	71.90	72.50	72.50
LR	72.20	72.22	74.30	74.91	71.10	71.20	71.20	71.30	71.50	71.50
MLP	70.40	90.90	89.60	89.60	88.40	88.70	91.50	91.91	92.10	92.60
RF	95.30	95.3	92.00	92.00	97.20	98.22	97.50	97.60	97.70	97.80
BFTree	89.40	89.50	87.60	88.01	90.20	90.40	90.10	90.20	90.12	90.13
BayesNet	77.20	77.60	76.60	76.60	76.80	76.80	77.10	77.10	77.30	77.30
AdaBoost	90.75	90.85	90.80	90.90	91.80	91.90	92.60	92.60	92.90	92.90
SVM	92.80	92.90	93.80	93.90	94.70	94.80	95.40	95.70	95.60	95.60

electrocardiogram data by section, were used to classify the normal elderly and the elderly with stroke based on machine learning. In this experiment, the performance was verified using the performance evaluation index defined in Section IV.B. Tables 5 and 6 summarize the experimental results using the performance indexes such as accuracy, F1-score, recall, and precision for each machine learning algorithm. In summary, the experiment was conducted using attributes from ECG and PPG bio-signals, which had not been used in previous studies. With 10-folder cross validation (CV), it can be seen that the C4.5, Multi-Layer Perceptron, Random Forest, AdaBoost, and SVM algorithms all show stable classification performance over 90%. The RandomForest method showed the highest classification accuracy of up to 98.31% in the 20-folder CV with the proper hyperparameter setting. In the second experiment, by applying Equation (3) defined in Section 3.C, only 12 out of 29 attributes were selected, and the selected optimal attributes were used. The selected 12 attributes are the set of {ECG_a_P, ECG_Q_R, ECG_R_S, ECG_PR_width, ECG_S_T, ECG_R_peak, ECG_S_peak, ECG_RRI, PPG_L_b_peak, PPG_R_a_b, PPG_R_b_d, and PPG_R_b_peak}. The selected 12 attributes consist of eight attributes related to ECG and four attributes related to PPG. Tables 7 and 8 summarized the performance evaluation with F1-score, recall, and precision for each algorithm with 12 optimal attributes. According to the experimental results, the

C4.5 decision tree, Multi-Layer Perceptron, Random Forest, BFTree, AdaBoost, and SVM algorithms showed more than 90% prediction accuracy with 10-folder CV. In particular, C4.5 Decision Tree shows the classification accuracy of 92.75% while RandomForest shows the highest classification accuracy of 97.71% in 20-folder CV with only 12 optimal property sets.

D. PREDICTION RESULT AND ANALYSIS OF REAL-TIME ELDERLY STROKE DISEASE PREDICTION BASED ON DEEP LEARNING

In this section, we show that it is possible to distinguish between stroke and normal elderly based on long short-term memory (LSTM) using real-time ECG and PPG multi-bio signals, which is good for time series analysis in the deep learning field. We trained LSTM [16], [52], [53], Bidirectional LSTM [52], [53], CNN-LSTM [56], [57], and CNN-Bidirectional LSTM [58], [59] using raw data of ECG and PPG to generate stroke prediction model. Bio-signals such as ECG and PPG are time series data that have sequential values according to time, and time information must be considered in the training process.

Therefore, in this paper, we modified and optimized the structure of a deep learning model suitable for time series for stroke disease prediction and used it in experiments [13], [14]. Recurrent neural networks (RNN) have been actively used for automatic translation, image caption generation, and

TABLE 9. Performance evaluation based on deep learning using ECG and PPG raw data (%).

Measures Methods	Accuracy	F1-score	Recall	Precision
LSTM	89.10	89.04	89.13	89.21
BiLSTM [1]	94.73	94.91	94.74	94.72
CNN-LSTM	99.15	99.11	99.13	99.22
CNN-BiLSTM [2]	96.92	96.91	96.90	96.90

[1] BiLSTM: Bidirectional LSTM, [2] CNN-BiLSTM: CNN-Bidirectional LSTM.

speech recognition, because the information inside the neural network persists [16], [52]. However, the RNN has a disadvantage in that, when the network is deep, the gradient value is lost and the learning performance is deteriorated; this is called the gradient vanishing problem. To solve this problem, an LSTM model that adds a cell state to the hidden state of RNN has been newly proposed [52], [53]. The LSTM consists of a forget gate, an input gate, and an output gate. Since the LSTM predicts the next state by considering the information before and after the current network layer, it can solve the long-term memory dependency problem of RNNs. Bidirectional RNN is a model that, to obtain meaningful results, not only infers in the forward direction of time series data, but also infers in the backward direction, from the future to the past [60]. That is, it is possible to predict a label for the current data through a past time series sequence and simultaneously predict a label for the current data through a future sequence. This model consists of a hidden layer with forward state information and a hidden layer with backward state information, and the two layers are not connected to each other. The input value is transmitted to both hidden layers, and the output layer can also receive values from the two hidden layers and calculate the final output value. To solve the problem of the long-term memory dependence on RNNs in bidirectional RNNs, the bidirectional LSTM model applied with the LSTM network was used in the experiment [54], [55]. CNN [29], [35], [56] is a specialized model for image classification and prediction based on a complex nonlinear model. Although LSTM shows good performance based on sophisticated time-series data, it has a problem in that the accuracy is deteriorated as the predicted values converge in one direction when there is no particular trend and when the data has various changes [56], [57]. Therefore, in this paper, to solve this problem, we used a CNN-LSTM structure that combines a CNN for extracting features from ECG and PPG bio-signals and an LSTM model that shows good performance in predicting the next step in time series data [54], [55]. Important features were extracted from the time series data of the input ECG and PPG bio-signals, and these values were passed through the LSTM model using past and future information to accurately predict the prognostic symptoms of stroke disease. Finally, CNN-Bidirectional LSTM extracts features from ECG and PPG bio-signals by combining the CNN model in front of the LSTM model [58], [59]. Next, a forward LSTM model that can predict from a past time point to a future time point and

a bidirectional LSTM model that can simultaneously predict a backward direction from a future time point are combined with CNN and used in the experiment. The ECG and PPG multi-modal bio-signal data are time series data, and the experimental results obtained using the deep learning models are summarized in Table 9. To ensure the objectivity of the experimental results, in the same way as in Tables 5 and 6, training and testing data are divided into 70% and 30%, 80% and 20%, 5-fold CV, 10-fold CV, and 20-fold CV. The performance indicators in Table 9 specify the average values of accuracy, F1-score, recall, and precision tested for each data set.

Table 9 presents the experimental results obtained by applying four deep learning models. The experimental result using the CNN-LSTM model showed the highest prediction accuracy of 99.15%. In particular, satisfactory experimental results were showed in the performance indices of F1-score, recall, and precision. The hyperparameter settings of the CNN-LSTM model with the best prediction performance are as follows: The epoch is set to 1,000, the batch size is set to 128, the learning rate is set to 0.003, the dropout is set to 0.5, and Adam is used as an optimizer. Bidirectional LSTM, CNN-LSTM, and CNN-BiLSTM models achieved more than 90% in both prediction accuracy and F1-score performance indexes. The deep learning prediction model can accurately predict stroke disease without the need for separate preprocessing or attribute extraction for each bio-signal. It was also experimentally verified that can be used in a healthcare service that notifies the elderly or their guardians in real-time, such as an emergency alarm.

V. SEMANTIC ANALYSIS AND DISCUSSION
A. SEMANTIC INTERPRETATION AND ANALYSIS OF STROKE DISEASE PREDICTION
In this section, we provide semantic analysis and interpretation of stroke disease in the elderly based on the C4.5 Decision Tree algorithm, which is a representative classification and prediction model of data mining and machine learning. The C4.5 decision tree algorithm is a white-box approach, and it enables semantic interpretation by analyzing the rules for the operation principle for the prediction of stroke disease in the elderly in detail. Semantic analysis was conducted based on the experimental results using the optimal

TABLE 10. The rules for stroke prediction with only 12 attributes of ECG and PPG bio signals (Figure 7).

| Rule #	The rule and in-depth analysis
1	IF ECG_S_peak ≤ -0.647736 and ECG_R_peak ≤ 0.978241 then Normal.
2	IF ECG_S_peak ≤ -0.647736 and ECG_R_peak > 1.43173 then Normal.
3	IF ECG_S_peak ≤ -0.647736 and $0.978241 <$ ECG_R_peak ≤ 1.43173 and ECG_Q_R ≤ 0.03 then Normal.
4	IF ECG_S_peak ≤ -0.647736 and $0.978241 <$ ECG_R_peak ≤ 1.43173 and ECG_Q_R > 0.03 then Stroke.
5	IF ECG_S_peak > -0.647736 and ECG_R_S ≤ 0.031 and ECG_R_peak > 0.572205 then Normal.
6	IF ECG_S_peak > -0.647736 and ECG_R_S ≤ 0.031 and ECG_R_peak ≤ 0.572205 and ECG_RRI > 0.852 then Stroke.
7	IF ECG_S_peak > -0.647736 and ECG_R_S ≤ 0.031 and ECG_R_peak ≤ 0.572205 and ECG_RRI ≤ 0.852 and PPG_R_b_peak > 0.179443 then Normal.
8	IF ECG_S_peak > -0.647736 and ECG_R_S ≤ 0.031 and ECG_R_peak ≤ 0.572205 and ECG_RRI ≤ 0.852 and PPG_R_b_peak ≤ 0.179443 and ECG_S_peak > -0.348206 then Normal.
9	IF ECG_S_peak > -0.647736 and ECG_R_S ≤ 0.031 and ECG_R_peak ≤ 0.572205 and ECG_RRI ≤ 0.852 and PPG_R_b_peak ≤ 0.179443 and ECG_S_peak ≤ -0.348206 then Stroke.
10	IF ECG_S_peak > -0.647736 and ECG_R_S > 0.031 and ECG_S_peak ≤ -0.329742 and ECG_R_peak ≤ 1.40976 and PPG_L_b_peak ≤ 0.341187 then Normal.
11	IF ECG_S_peak > -0.647736 and ECG_R_S > 0.031 and ECG_S_peak ≤ -0.329742 and $0.899658 <$ ECG_R_peak ≤ 1.40976 and PPG_L_b_peak ≤ 0.341187 then Normal.
12	IF ECG_S_peak > -0.647736 and ECG_R_S > 0.031 and ECG_S_peak ≤ -0.329742 and ECG_R_peak ≤ 0.899658 and PPG_L_b_peak > 0.341187 and ECG_RRI ≤ 0.693 then Stroke.
13	IF ECG_S_peak > -0.647736 and ECG_R_S > 0.031 and ECG_S_peak ≤ -0.329742 and ECG_R_peak ≤ 0.899658 and PPG_L_b_peak > 0.341187 and ECG_RRI > 0.693 then Normal.
14	IF ECG_S_peak > -0.647736 and $0.031 <$ ECG_R_S ≤ 0.041 and ECG_S_peak ≤ -0.329742 and ECG_R_peak > 1.40976 then Stroke.
15	IF ECG_S_peak > -0.647736 and ECG_R_S > 0.041 and ECG_S_peak ≤ -0.329742 and ECG_R_peak > 1.40976 then Normal.
16	IF ECG_S_peak > -0.647736 and ECG_R_S > 0.031 and ECG_S_peak > -0.329742 and ECG_PR_width ≤ 0.137 then Stroke.
17	IF ECG_S_peak > -0.647736 and ECG_R_S > 0.031 and ECG_S_peak > -0.329742 and ECG_PR_width > 0.225 then Stroke.
18	IF ECG_S_peak > -0.647736 and $0.031 <$ ECG_R_S ≤ 0.038 and ECG_S_peak > -0.329742 and $0.137 <$ ECG_PR_width ≤ 0.225 then Normal.
19	IF ECG_S_peak > -0.647736 and ECG_R_S > 0.038 and ECG_S_peak > -0.329742 and $0.137 <$ ECG_PR_width ≤ 0.225 then Stroke.
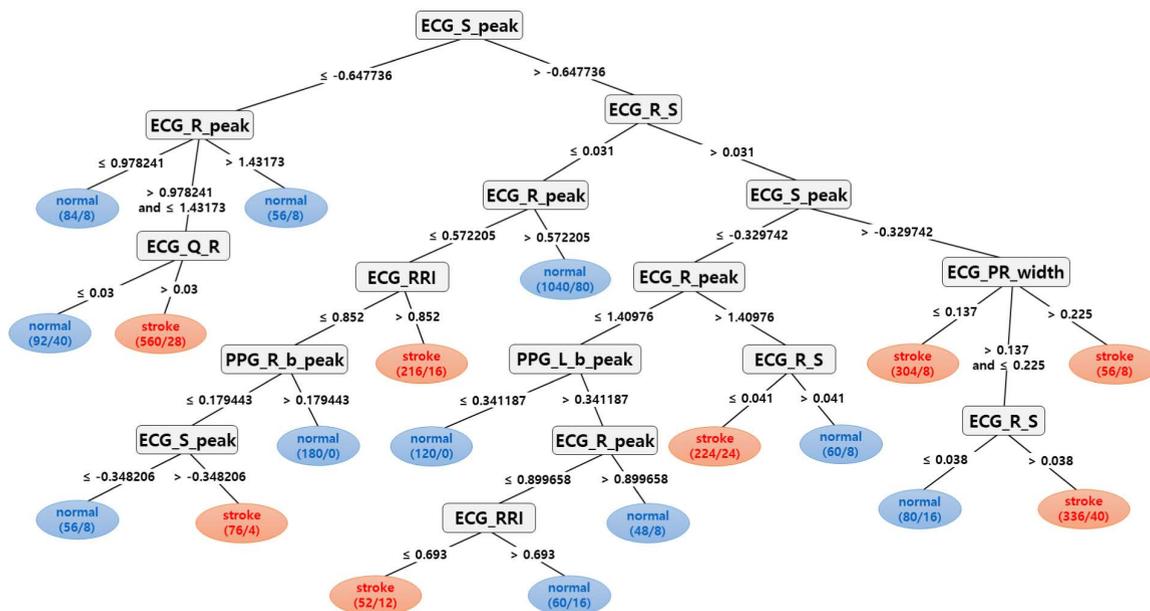
FIGURE 7. The C4.5 decision tree for stroke prediction based on 12 attributes of ECG and PPG (19 rules).

subset selection method that only involved 12 attributes, as listed in Table 7 of Section VI.B. There were 287 normal and 287 stroke elderly patients, respectively, and raw data were obtained by dividing bio-signals for each subject in

units of 5 seconds. A data set was constructed by repeating data extraction a total of 5 times for each subject. Data including noise values or missing values were deleted through verification by the medical staff. In this study, 2,520 data instances were extracted for each normal and stroke elderly patients classes. A predictive model was trained by using 2,016 instances randomly selected by 80% of data instances for each class, and the test was conducted with the remaining 504 data instances that did not participate in training. Table 7 shows that the C4.5 decision tree algorithm confirmed the prediction accuracy of 91.08% and a stable performance of 91.10% in recall and precision. Figure 7 shows the decision tree used to predict stroke disease in the elderly with only 12 attributes. When constructing a decision tree, normal and stroke elderly patients can be accurately classified and predicted using only 12 of the 29 attributes defined in this system (see Table 1). Here, the numerical values in the leaf node are models trained only with the learning data of the normal elderly and stroke patients, 80% of which are randomly extracted. The number of leaf nodes means the number correctly predicted by the classifier and the number misclassified. The accuracy of the training model of the C4.5 decision tree was 91.76%. Only eight of the 12 attributes {ECG_S_peak, ECG_R_peak, ECG_Q_R, ECG_R_S, ECG_RRI, PPG_R_b_peak, PPG_L_b_peak, and ECG_PR_width} were used in the generated decision tree as shown in Figure 7. Nineteen rules can be obtained from Figure 7, and each rule is listed and semantically analyzed in detail in Table 10. Among the attributes of ECG and PPG bio-signals, the S_peak value of ECG is the most important attribute, and it is located at the top node of the decision tree. ECG_R_peak and ECG_R_S are also very significant attributes for predicting stroke. In generating a stroke disease prediction model using multi-modal bio-signals of ECG and PPG, the results verified that accuracy prediction is possible even when using only eight out of 12 attributes (six ECG attributes and two PPG attributes). According to Rule 4 of Table 10, when the S_peak value of the ECG is small but the Q_R value is 0.03ms or more, the probability of predicting a stroke patient is analyzed to be high. This is used by medical staff to determine abnormalities in cardiovascular disease. It is consistent with the occurrence of symptoms of Q_R being irregular or longer than that of normal elderly. According to Rule 6, the RRI value of ECG above 0.852ms is also associated with a slower and irregular heart rate than normal elderly, which is suspected to be bradyarrhythmia or cardiovascular disease. These symptoms are interpreted to be caused by abnormalities of the autonomic nerves and sympathetic nerves due to pre-stroke, and they can be reflected in the RRI values of the ECG and PPG bio-signals measured in real-time. As a result, the prediction accuracy of 91.08% of stroke prognostic symptoms in the elderly was verified using only 19 rules. In addition, as presented in Table 10, the utilization value of services in hospitals and medical institutions is high due to the rule-based semantic interpretation.

VI. CONCLUSION

In this paper, we propose a system that provides the semantic analysis of diseases in the elderly using multiple biological signals of ECG and PPG collected from walking during the daily life of the elderly. The proposed system collects multiple bio-signals of ECG and PPG in real-time, and it can immediately detect and predict prognostic symptoms of stroke disease in the elderly. A machine learning-based prediction model study was performed using multiple bio-signal data, which involves dividing the signal waveform into specific sections, and relatively accurate prediction results and semantic interpretations were obtained using this model. In this paper, using the proposed attributes, it was experimentally verified that the prognostic symptoms of stroke patients can be accurately predicted by more than 90% based solely on ECG and PPG collected while walking. To summarize the experimental and verification results, we confirmed that we can accurately predict 91.56% C4.5 Decision Tree, 97.51% RandomForest, and 99.15% CNN-LSTM models for deep learning, by separating stroke and general elderly into 10-folder CV datasets. The system proposed in this paper has great academic value in that it can accurately predict the prognostic symptoms and onset of stroke by measuring ECG and PPG at low cost which can be worn with little inconvenience during daily life. Various bio-signal data collected in daily life can provide objective interpretation information to stroke patients or medical staff with a high recurrence rate. The experimental results verified that this method can be used for practical healthcare services which reduce the aftereffects of stroke and prevent emergency situations through constant monitoring.

In future studies, we will conduct in-depth analysis and predictive experiments of stroke disease by analyzing various bio-signals such as EEG, EMG, foot pressure, and motion, as well as electronic medical records (EMRs) and MRI image information.

REFERENCES

[1] S. De Raedt, A. De Vos, and J. De Keyser, "Autonomic dysfunction in acute ischemic stroke: An underexplored therapeutic area?" *J. Neurol. Sci.*, vol. 348, nos. 1–2, pp. 24–34, Jan. 2015.

[2] C. P. Warlow, "Epidemiology of stroke," *Lancet*, vol. 352, pp. 1–4, Oct. 1998.

[3] K.-D. Seo, M. J. Kang, G. S. Kim, J. H. Lee, S. H. Suh, and K.-Y. Lee, "National trends in clinical outcomes of endovascular therapy for ischemic stroke in South Korea between 2008 and 2016," *J. Stroke*, vol. 22, no. 3, pp. 412–415, Sep. 2020.

[4] T. D. Musuka, S. B. Wilton, M. Traboulsi, and M. D. Hill, "Diagnosis and management of acute ischemic stroke: Speed is critical," *Can. Med. Assoc. J.*, vol. 187, no. 12, pp. 887–893, Sep. 2015.

[5] Q. Song, X. Liu, W. Zhou, L. Wang, X. Zheng, X. Wang, and S. Wu, "Long sleep duration and risk of ischemic stroke and hemorrhagic stroke: The Kailuan prospective study," *Sci. Rep.*, vol. 6, no. 1, pp. 1–9, Sep. 2016.

[6] J. Yu, S. Park, H. Lee, C. S. Pyo, and Y. S. Lee, "An elderly health monitoring system using machine learning and in-depth analysis techniques on the NIH stroke scale," *Mathematics*, vol. 8, no. 7, pp. 1–16, Jul. 2020.

[7] World Health Organization. *The Top 10 Causes of Death*. Accessed: Apr. 22, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

[8] *UNDESA World Social Report 2020*. Accessed: Apr. 22, 2022. [Online]. Available: https://www.un.org/development/desa/dspd/world-social-report/2020-2.html

[9] S. Park, M. J. Yang, S. N. Ha, and J. S. Lee, "Effective anti-aging strategies in an era of super-aging," *J. Menopausal Med.*, vol. 20, no. 3, pp. 85–89, Dec. 2014.

[10] B. C. Meyer and P. D. Lyden, "The modified national institutes of health stroke scale: Its time has come," *Int. J. Stroke*, vol. 4, no. 4, pp. 267–273, Aug. 2009.

[11] D. Schlegel, S. J. Kolb, J. M. Luciano, J. M. Tovar, B. L. Cucchiara, D. S. Liebeskind, and S. E. Kasner, "Utility of the NIH stroke scale as a predictor of hospital disposition," *Stroke*, vol. 34, no. 1, pp. 134–137, Jan. 2003.

[12] J. Yu, D. Kim, H. Park, S. Chon, K. H. Cho, S. Kim, S. Yu, S. Park, and S. Hong, "Semantic analysis of NIH stroke scale using machine learning techniques," in *Proc. PlatCon*, Jeju, South Korea, Jan. 2019, pp. 82–86.

[13] J. Yu, S. Park, S. H. Kwon, C. M. B. Ho, C. S. Pyo, and H. Lee, "AI-based stroke disease prediction system using real-time electromyography signals," *Appl. Sci.*, vol. 10, no. 19, pp. 1–19, Sep. 2020.

[14] Y. A. Choi, S. Park, J. A. Jun, C. M. B. Ho, C. S. Pyo, H. Lee, and J. Yu, "Machine-learning-based elderly stroke monitoring system using electroencephalography vital signals," *Appl. Sci.*, vol. 11, no. 4, pp. 1–18, Feb. 2021.

[15] M. Lee, J. Ryu, and D. Kim, "Automated epileptic seizure waveform detection method based on the feature of the mean slope of wavelet coefficient counts using a hidden Markov model and EEG signals," *ETRI J.*, vol. 42, no. 2, pp. 217–229, Apr. 2020.

[16] Y. A. Choi, S. J. Park, J. A. Jun, C. S. Pyo, K. H. Cho, H. S. Lee, and J. H. Yu, "Deep learning-based stroke disease prediction system using real-time bio signals," *Sensors*, vol. 21, no. 13, pp. 1–17, Jun. 2021.

[17] S.-M. Lai, P. W. Duncan, and J. Keighley, "Prediction of functional outcome after stroke: Comparison of the orpington prognostic scale and the NIH stroke scale," *Stroke*, vol. 29, no. 9, pp. 1838–1842, Sep. 1998.

[18] S. Shrestha, R. S. Poudel, D. Khatiwada, and L. Thapa, "Stroke subtype, age, and baseline NIHSS score predict ischemic stroke outcomes at 3 months: A preliminary study from Central Nepal," *J. Multidisciplinary Healthcare*, vol. 8, pp. 443–448, Oct. 2015.

[19] E. A. Sartor, K. Albright, A. K. Boehme, M. M. Morales, A. Shaban, J. C. Grotta, S. I. Savitz, and S. Martin, "The NIHSS score and its components can predict cortical stroke," *J. Neurol. Disorders Stroke*, vol. 2, no. 1, pp. 73–78, Sep. 2013.

[20] R. G. Hart and J. L. Halperin, "Atrial fibrillation and stroke: Concepts and controversies," *Stroke*, vol. 32, no. 3, pp. 803–808, Mar. 2001.

[21] G. Y. H. Lip and H. S. Lim, "Atrial fibrillation and stroke prevention," *Lancet Neurol.*, vol. 6, no. 11, pp. 981–993, 2007.

[22] J. A. Reiffel, "Atrial fibrillation and stroke: Epidemiology," *Amer. J. Med.*, vol. 127, no. 4, pp. 15–16, Apr. 2014.

[23] P. A. Wolf, R. B. D'Agostino, A. J. Belanger, and W. B. Kannel, "Probability of stroke: A risk profile from the Framingham study," *Stroke*, vol. 22, no. 3, pp. 312–318, Mar. 1991.

[24] S. H. Jee, I. Suh, I. S. Kim, and L. J. Appel, "Smoking and atherosclerotic cardiovascular disease in men with low levels of serum cholesterol: The Korea medical insurance corporation study," *Jama*, vol. 282, no. 22, pp. 2149–2155, Dec. 1999.

[25] D. B. Geselowitz, "On the theory of the electrocardiogram," *Proc. IEEE*, vol. 77, no. 6, pp. 857–876, Jun. 1989.

[26] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 289–294, Mar. 2003.

[27] F. Rundo, S. Conoci, A. Ortis, and S. Battiato, "An advanced bio-inspired photoplethysmography (PPG) and ECG pattern recognition system for medical assessment," *Sensors*, vol. 18, no. 2, pp. 1–22, Jan. 2018.

[28] R. Couceiro, P. Carvalho, J. Henriques, M. Antunes, M. Harris, and J. Habetha, "Detection of atrial fibrillation using model-based ECG analysis," in *Proc. 19th Int. Conf. Pattern Recognit.*, Tampa, FL, USA, Dec. 2008, pp. 1–5.

[29] Z. Xiong, M. K. Stiles, and J. Zhao, "Robust ECG signal classification for detection of atrial fibrillation using a novel neural network," in *Proc. Comput. Cardiol. (CinC)*, Rennes, France, vol. 44, Sep. 2017, pp. 1–4.

[30] J. M. Bumgarner, C. T. Lambert, A. A. Hussein, D. J. Cantillon, B. Baranowski, K. Wolski, and K. G. Tarakji, "Smartwatch algorithm for automated detection of atrial fibrillation," *J. Amer. College Cardiol.*, vol. 71, no. 21, pp. 2381–2388, Mar. 2018.

[31] S. Babaeizadeh, R. E. Gregg, E. D. Helfenbein, J. M. Lindauer, and S. H. Zhou, "Improvements in atrial fibrillation detection for real-time monitoring," *J. Electrocardiol.*, vol. 42, no. 6, pp. 522–526, Nov. 2009.

[32] G. Sannino and G. De Pietro, "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection," *Future Gener. Comput. Syst.*, vol. 86, pp. 446–455, Sep. 2018.

[33] E. J. D. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Comput. Methods Programs Biomed.*, vol. 127, pp. 144–164, Apr. 2016.

[34] J. S. Wang, W. C. Chiang, Y. L. Hsu, and Y. T. C. Yang, "ECG arrhythmia classification using a probabilistic neural network with a feature reduction method," *Neurocomputing*, vol. 116, pp. 38–45, Sep. 2013.

[35] Ö. Yıldırım, P. Plawiak, R.-S. Tan, and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ECG signals," *Comput. Biol. Med.*, vol. 102, no. 1, pp. 411–420, Nov. 2018.

[36] NIH. *National Heart, Lung, and Blood Institute*. Accessed: Apr. 22, 2022. [Online]. Available: https://www.nhlbi.nih.gov/health/heart-failure/causes

[37] W. Nichols, "Clinical measurement of arterial stiffness obtained from noninvasive pressure waveforms," *Amer. J. Hypertension*, vol. 18, no. 1, pp. 3–10, Jan. 2005.

[38] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, pp. 1–39, Feb. 2007.

[39] J. J. Oliver and D. J. Webb, "Noninvasive assessment of arterial stiffness and risk of atherosclerotic events," *Arteriosclerosis, Thrombosis, Vascular Biol.*, vol. 23, no. 4, pp. 554–566, Feb. 2003.

[40] D. Jarchi and A. J. Casson, "Towards photoplethysmography-based estimation of instantaneous heart rate during physical activity," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2042–2053, Sep. 2017.

[41] *MIT-BIH Arrhythmia Database*. Accessed: Apr. 22, 2022. [Online]. Available: https://physionet.org/content/mitdb/1.0.0/

[42] *Weka, Machine Learning Software in Java*. Accessed: Apr. 22, 2022. [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/

[43] H. Kamel, E. Z. Soliman, S. R. Heckbert, R. A. Kronmal, W. T. Longstreth, S. Nazarian, and P. M. Okin, "P-wave morphology and the risk of incident ischemic stroke in the multi-ethnic study of atherosclerosis," *Stroke*, vol. 45, no. 9, pp. 2786–2788, Sep. 2014.

[44] J. He, G. Tse, P. Korantzopoulos, K. P. Letsas, S. Ali-Hasan-Al-Saegh, H. Kamel, G. Li, G. Y. H. Lip, and T. Liu, "P-wave indices and risk of ischemic stroke: A systematic review and meta-analysis," *Stroke*, vol. 48, no. 8, pp. 2066–2072, Aug. 2017.

[45] H. Christensen, A. F. Christensen, and G. Boysen, "Abnormalities on ECG and telemetry predict stroke outcome at 3 months," *J. Neurol. Sci.*, vol. 234, nos. 1–2, pp. 99–103, Jul. 2005.

[46] A. Schuchert, G. Behrens, and T. Meinertz, "Impact of long-term ECG recording on the detection of paroxysmal atrial fibrillation in patients after an acute ischemic stroke," *Pacing Clin. Electrophysiol.*, vol. 22, no. 7, pp. 1082–1084, Jul. 1999.

[47] D. Jabaudon, J. Sztajzel, K. Sievert, T. Landis, and R. Sztajzel, "Usefulness of ambulatory 7-day ECG monitoring for the detection of atrial fibrillation and flutter after acute stroke and transient ischemic attack," *Stroke*, vol. 35, no. 7, pp. 1647–1651, Jul. 2004.

[48] I.-S. Oh, J.-S. Lee, and B.-R. Moon, "Hybrid genetic algorithms for feature selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1424–1437, Nov. 2004.

[49] M. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, Dept. Comput. Sci., Univ. Waikato, Hamilton, NZ, USA, 1998.

[50] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 3rd ed. Waltham, MA, USA, Morgan Kaufmann, 2011, pp. 322–325.

[51] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: An overview," 2020, *arXiv:2008.05756*.

[52] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[53] X. Xiao, S. Zhang, F. Mercaldo, G. Hu, and A. K. Sangaiah, "Android malware detection based on system call sequences and LSTM," *Multimedia Tools Appl.*, vol. 78, no. 4, pp. 3979–3999, 2019.

[54] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[55] A. Graves, S. Fernández, and J. Schmidhuber, "Bidirectional LSTM networks for improved phoneme classification and recognition," in *Proc. ICANN*, Warsaw, Poland, Sep. 2005, pp. 799–804.

[56] J. Wang, L.-C. Yu, K. R. Lai, and X. Zhang, "Dimensional sentiment analysis using a regional CNN-LSTM model," in *Proc. Annu. Meeting ACL*, Berlin, Germany, vol. 2, Aug. 2016, pp. 225–230.

[57] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, no. 1, pp. 72–81, Sep. 2019.

[58] J.-H. Jeong, K.-H. Shim, D.-J. Kim, and S.-W. Lee, "Brain-controlled robotic arm system based on multi-directional CNN-BiLSTM network using EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1226–1238, May 2020.

[59] Z.-X. Liu, D.-G. Zhang, G.-Z. Luo, M. Lian, and B. Liu, "A new method of emotional analysis based on CNN-BiLSTM hybrid neural network," *Cluster Comput.*, vol. 23, no. 4, pp. 2901–2913, Dec. 2020.

[60] S. Zhang, D. Zheng, X. Hu, and M. Yang, "Bidirectional long short-term memory networks for relation classification," in *Proc. PACLIC*, Shanghai, China, Oct./Nov. 2015, pp. 73–78.

JAEHAK YU received the B.S. degree in computer science from Konkuk University, Republic of Korea, in 2001, and the M.S. and Ph.D. degrees in computer science from Korea University, Republic of Korea, in 2003 and 2010, respectively.

Since 2010, he has been with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea, where he is currently a Senior Member with the Department of Knowledge-Converged Super Brain (KSB) Convergence Research. His recent research interests include artificial intelligence (AI), machine learning, deep learning, data analysis and engineering, the IoT/Internet of Everything (IoE) platform, and intelligent cyber/network security.

SEJIN PARK received the Ph.D. degree in industrial engineering from Korea University, in 1994.

Since 1988, he has been with the Korea Research Institute of Standards and Science (KRISS), where he served in various positions, including the Director of Convergence Technology and the Head of Ergonomics Research and Medical Metrology. He is currently the Director of the Data Center for Korean Body Measurement, Korea Research Institute of Standards and Science (KRISS), supported by Ministry of Knowledge Economy, South Korea. He is also a Professor with the Department of Medical Physics, University of Science and Technology (UST), South Korea. His research interests include human factors and ergonomics, biomechanics, emotion and sensibility, human vibration and seating comfort, human–computer interaction (HCI), the Internet of Everything (IoE), gerontechnology, and anthropometry. He served as the President for the Korean Society of Emotion and Sensibility (KOSES) and the Ergonomics Society of Korea.

SOON-HYUN KWON received the B.S. degree in computer science from Incheon National University, Republic of Korea, in 1998, and the M.S. degree in computer science from Soongsil University, Republic of Korea, in 2003. From 2006 to 2009, he was a member of Engineering Staff at the Ontology Reasoner Development Team, where he worked on ontology inference algorithm and development. From 2010 to 2015, he was a Senior Member of Engineering Staff at the IoT Platform Research Section, where he has developed the Semantic IoT service platform. Since 2013, he has been with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea, where he is currently a Senior Member with the Knowledge-Converged Super Brain (KSB) Convergence Research Department. His research interests include semantic web, linked data, ontology inference, artificial intelligence (AI), and knowledge augmenting.

KANG-HEE CHO received the Ph.D. degree in rehabilitation medicine from the Graduate School, Chungnam National University. He is currently a Professor with the Department of Rehabilitation Medicine, College of Medicine, Chungnam National University. His work focuses specifically on rehabilitation medicine, rehabilitation medicine devices, and medical engineering. Recently, he founded the 'Data Center for Korean Range of Motion of Joint' and collect data on the Joint range of motion of Korean.

HANSUNG LEE received the B.S., M.S., and Ph.D. degrees in computer science from Korea University, Republic of Korea, in 1996, 2002, and 2008, respectively. From July 1996 to July 1999, he worked at Daewoo Engineering Company, Republic of Korea. From November 2009 to November 2014, he worked with the Electronics and Telecommunications Research Institute (ETRI), Republic of Korea, as a Senior Member of the Research Staff. From December 2014 to August 2019, he worked with Samsung Electronics Company Ltd., Republic of Korea. From September 2019 to February 2021, he was an Assistant Professor with Youngsan University, Republic of Korea. In March 2021, he joined Andong National University, Republic of Korea. His current research interests include machine/deep learning, computer vision, computer/network security, optimization, data mining, and big-data analytics.

● ● ●