

Received August 16, 2021, accepted August 31, 2021, date of publication September 1, 2021, date of current version September 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3109806

Big-ECG: Cardiographic Predictive Cyber-Physical System for Stroke Management

IQRAM HUSSAIN^{ID}, (Graduate Student Member, IEEE), AND SE JIN PARK^{ID}

Center for Medical Convergence Metrology, Korea Research Institute of Standards and Science, Daejeon 34113, South Korea

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

Department of Medical Physics, University of Science and Technology, Daejeon 34113, South Korea

Corresponding author: Se Jin Park (sjpark@kriss.re.kr)

This work was supported by the National Research Council of Science and Technology (NST) Grant through Korean Government former Ministry of Science, ICT and Future Planning (MSIP), succeeded by Ministry of Science and ICT, under Grant CRC-15-05-ETRI.

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of the Korea Research Institute of Standards and Science, Daejeon, South Korea under Application No. KRIS-IRB-2016-05-19, and performed under the guidelines of the Declaration of Helsinki.

ABSTRACT Electrocardiogram (ECG) is sensitive to autonomic dysfunction and cardiac complications derived from ischemic or hemorrhage stroke and is supposed to be a potential prognostic tool in stroke identification and post-stroke treatment. ECG data generated cannot be real-time accumulated, processed, and used for enterprise-level healthcare and wellness services with the existing cardiovascular monitoring system used in hospitals. This study aims to assess the feasibility of a cyber-physical cardiac monitoring system to classify stroke patients with altered cardiac activity and healthy adults. Here, we propose Big-ECG, a cyber-physical cardiac monitoring system for stroke management, consisting of a wearable ECG sensor, data storage and data analysis in a big data platform, and health advisory services using data analytics and medical ontology. We investigated our proposed ECG-based patient monitoring system with 45 stroke patients (average age 70.8 years old, 68% men) admitted to the rehabilitation center of the hospital and 40 healthy elderly volunteers (average age 75.4 years old, 38% men). We recorded ECG at resting state using a single-channel ECG patch within three months of diagnosis of ischemic stroke (clinically confirmed). In statistical results, ECG fiducial features, RR-I, QRS, QT, ST, and heart rate variability (HRV) features, SDDSD, LF/HF, LF/(LF + HF), and HF/(LF + HF) are observed as significantly distinctive biomarkers for the stroke group relative to the healthy control group. The Random Trees model presented the best classification performance (overall accuracy: 95.6%) utilizing ECG fiducial variables. This system may assist healthcare enterprises in prognosis and rehabilitation management during post-stroke treatment.

INDEX TERMS Cyber-physical systems, electrocardiography, biomedical monitoring, big data applications, biomedical informatics.

I. INTRODUCTION

Stroke, a primary neurovascular disease in adulthood, is the world's second leading cause of death in the elderly community [1]. Hemorrhagic events, such as a stroke, occurs due to the blood vessel's rupture in the brain and hamper the supply of oxygen to brain tissue at the lesion site causing brain cell death. This damage to the brain tissue affects the central nervous system. Furthermore, stroke is commonly

associated with autonomic dysfunction [2] and cardiovascular responses [3], which may increase mortality and morbidity rates. Early prediction of stroke symptoms affects mortality, rehabilitation, cost of post-stroke treatment, and quality of life [4]. Often, the stroke symptoms are not noticeable in the early stages of an ischemic event. Therefore, the decision to refer a stroke survivor to a clinical diagnostic center for brain imaging and pathological evaluation may delay. Late diagnostics of ischemic stroke can lead to motor impairment, sensory impairment, cognitive impairment, and even death. The prospect of improvement after stroke differs with the

The associate editor coordinating the review of this manuscript and approving it for publication was Humaira Nisar^{ID}.

severity of the initial motor and cognitive deficit. The economic burden of post-stroke treatment is among the fastest-growing expenses for healthcare [5].

Tracking the physiological signals is one of the essential methods for disease prognostics and clinical management. Stroke is a neurological disease, and electroencephalography (EEG) is a useful tool for early prognostics of stroke [6]. Besides, ischemic stroke affects the autonomic nervous system (ANS), cardiovascular activity. As electrocardiography (ECG) is the representative physiological signal of cardiovascular health and the autonomic nervous system, cardiac monitoring is one of the keys for stroke prediction. Several ECG studies have been reported the quantitative ECG measurements in clinical applications to evaluate the relationship between cardiac, neurological, and functional outcomes of ischemic stroke [2]. Changes in the ECG-derived heart rate variability (HRV) are the biomarker of the sympathetic and parasympathetic activity of the ANS regulating most visceral and metabolic processes. Cardiac dynamics can be tracked using nonlinear HRV measures and studied to predict sleep apnea [7] and congestive heart failure [8]. Ischemic stroke impairs autonomic function, characterized by a dominance of sympathetic activity. Cardiac abnormality, such as myocardial ischemia, is associated with stroke patients. The most common ECG fiducial changes include depressed ST-segments, prolonged QT-interval, flat or inverted T-waves, and U-waves [9].

For a detailed diagnostics of stroke, including identification of the stroke lesion on the brain, and evaluation of lesion size and location, computed tomography (CT) and magnetic resonance imaging (MRI) is the most useful tool to understand the anatomy of the brain and to determine the scope of diagnosis for both types of stroke (thrombosis or hemorrhage) [10]. Continuous monitoring of high-risk patients with a history of acute stroke or transient ischemic attack (mini stroke) using CT and MRI is impractical [6]. For the prognosis of Stroke, ECG changes can be useful in daily life and the clinical environment [11], [12]. Moreover, ECG or vital sign functionality is present in most fitness trackers and wellness devices. Real-time tracking of heart activity is an affordable and effective way to predict high-risk stroke patients' cardiovascular health status with underlying heart diseases.

With the advancement of a cyber-physical system, big data, and Healthcare 4.0 in medicine, a real-time biosignal-based patient monitoring system draws much attention. Elderly adults are most vulnerable to several life-threatening diseases, such as ischemic stroke, heart disease. Besides, government and healthcare agencies are looking for an innovative and effective way to manage senior citizens' treatment. Rehabilitation is an important step to recover from cardiovascular and neurological disorders. Healthcare providers generally record ECG using an existing standard 12-lead ECG system with multiple electrodes in medical centers and hospitals. These ECG studies also require trained medical staff and clinical settings. Besides, traditional

ECG methods use multiple spatial positions to measure the heart's electrical activity correctly. These kinds of long clinical preparation and expert skill demand can delay the prognosis of acute diseases. Moreover, cardiovascular and neurological impairment resulting from stroke increases the risk of cardiac morbidity and mortality during the post-stroke period [3]. So, the real-time cardiac monitoring system has achieved considerable interest for post-stroke rehabilitation management.

ECG data generated in healthcare centers cannot be real-time stored, transformed, and utilized for enterprise-level clinical and wellness services with the present cardiovascular monitoring system. Generally, doctors' intervention is required to interpret ECG for clinical decision-making. As the application of wearable medical devices is growing and home patient monitoring system is getting popularity during COVID-19 era, an automated ECG analytics platform may come to be an assistive tool for medical experts and patient caregivers. Not enough extensive studies investigated cardiac activity using portable ECG and cloud-based live processing for stroke patients' prognostics and rehabilitation. In addition, ECG-derived HRV parameters were studied earlier for machine learning based stroke prediction [9]. ECG fiducial feature based machine-learning approach was not clinically explored yet in the case of ischemic stroke. In summary, it requires a real-time or near-real-time ECG system for monitoring cardiovascular activity in a stroke patient's daily life setting. We proposed Big-ECG, capable of tracking the cardiac signal, analyzing data in the big data platform, and providing health analytics as a service. This system can generate alerts as feedback for the assistance of the emergency rescue services if stroke-predictive cardiac features exceed any lethal criteria. Big-ECG is a cyber-physical platform that combines clinical ECG and big data analytics.

We hypothesized that a portable ECG device would immediately detect cardiovascular activity. Data analytics based on biosignal processing, statistical analysis, and robust machine learning techniques will be consistent methods for predicting cardiac health during stroke onset and post-stroke rehabilitation.

We aim to develop the Big-ECG, a cyber-physical ECG system for stroke, cardiovascular disease prediction in daily life and clinical environments. The key contributions of this paper can be summarized as follows:

- We established a Big-ECG platform integrating the wearable ECG patch, data streaming to a cloud server, real-time signal processing with Hadoop and Spark ecosystem, live dashboards for the customers, doctors, and service managers for cloud-based prognostics of the ischemia and heart diseases.
- We discovered stroke-impaired ECG indices, including HRV measures, fiducial features using statistical analysis, and significantly important features through hypothesis tests.

- We utilized the machine-learning algorithms to categorize the ischemic stroke group and the healthy control group for acute stroke prediction.

We organized the remainder of this article into six sections. We narrated this article's technical background in Section II, exploring the state-of-art techniques of Healthcare 4.0 and big data. The proposed cyber-physical ECG-based health monitoring platform was described in Section III, followed by the datasets and the methodology used to validate the system's predictive capability. After that, the results are reported in Section V, trailed by the discussion. Lastly, we stated the conclusions are in Section VII.

II. BACKGROUND

A. CYBER-PHYSICAL SYSTEM AND HEALTHCARE 4.0

The Cyber-Physical System (CPS) is an innovative system with integrated computing and physical capabilities that enables new ways to communicate with humans [13]. The interaction between physical and digital elements has come to play an essential role in various domains. CPS, adapted in the industrial sector, is now being implemented in healthcare. Within a healthcare context, the use of CPS has led to Smart Healthcare. From the perspective of this new revolution, a vast quantity of CPS shapes current healthcare systems involving devices, technologies, solutions, and ventures. The vital components of these CPS are composed of a blend of enabling technologies, comprising smart medical devices, diagnostics process automation, autonomous robots, Internet of Things (IoT) devices, medical Big Data, Fog, and Cloud Computing [14]. CPS is the crucial technology of Healthcare 4.0 [15]. Healthcare 4.0 is a consecutive revolution of the complete healthcare system, including the intelligent manufacturing of medicine and medical devices, cyber-physical patient monitoring, health analytics, telemedicine, healthcare logistics, personalized and precision medicine, assisted living, and rehabilitation.

B. BIG DATA TECHNOLOGIES IN HEALTHCARE

Big data technologies are shaping the world; healthcare is no exception. The real-time transactional medical data and the accumulated historical electronic health records (EHR) data in a hospital are vital medical decisive tools for clinicians and other care providers for the patient's best care or services. Healthcare big data builds up with a massive volume of structured and unstructured medical records. Usually, the structured database resulted from several sources, such as patient demographics, living habits, diagnostic tests, and diseases. On the other hand, patients' medical history, doctor's interrogation records contribute to accumulating vast unstructured datasets. Hospitals are handling this Big medical Data acquisition, processing and analytics, storage, retrieving real-time data, and collecting historical medical data using suitable Big data technologies [16]. Big Data technologies include the low-cost open-source Hadoop ecosystem, Elasticsearch (ES), and the relational database (RDB) and

Hadoop-HadoopDB [17]. The Hadoop ecosystem comprises of Hadoop distributed file system (HDFS), MapReduce algorithm, and other analytical tools for handling, analyzing Big Data to make it mature and enterprise-ready.

C. HEALTHCARE WEARABLES AND BIOSIGNALS

With the rise of wearable healthcare and wellness devices, the source of healthcare data is expanding rapidly. High-speed network, wearable physiological devices are enabling smart homes to the part of the medical CPS. Several physiological signals are measured in the hospitals and stored as a numerical value or digital portable documents format (pdf). ECG gives information about cardiac activity by measuring the electrical behavior of the heart. Electromyography (EMG) shows the muscle's health by reading the bioelectrical activity generated by muscle fibers. Electroencephalogram (EEG) reveals the neurological status by measuring the electrical activity of the brain. Photoplethysmogram (PPG), galvanic skin response (GSR), electrooculography (EOG) are examples of biosignals showing the health status of the patients. Several human behavioral signals, motion, gait, and postural parameters portray physical and behavioral health. This kind of unstructured data is very complicated to analyze using a big data platform. Clinical decision-making using those records need a physician or doctor's assistance. The rapid advancement of miniature bio-signal processing hardware, application programming interface (API), communication technologies, and machine-learning techniques enable wearables healthcare devices to deal with real-time health analytics [18], [19]. Wearable health trackers can record biosignals in daily life setup, such as home activities, walking, driving, sleeping, and the widening era of the healthcare domain. Nowadays, a patient's real-time physiological data are generally acquired using wearable sensors, such as a vital sign tracker, an activity tracker [18], [4], [20], [21], and a sleep tracker [6], [22]. Wearable devices are utilized in wellness and health stimulation, such as vagus nerve stimulation, microwave brain stimulation [23].

III. BIG-ECG: A CYBER-PHYSICAL CARDIAC MONITORING SYSTEM

Big-ECG, a novel cardiac monitoring system, consists of a portable ECG sensor system, the data acquisition interface, the big data storage and processing, the knowledge-base, and the healthcare service dashboard, as shown in Figure 1. This section explored the proposed cyber-physical cardiac monitoring architecture for real-time monitoring in detail. First, the components of the system architecture are introduced (Section III-A). Finally, we will demonstrate the dataflow to assess the real-time cardiac tracking using the CPSs (Section III-B).

A. SYSTEM ARCHITECTURE

This architecture aims to process the physiological signal to monitor in real-time health status, providing expert

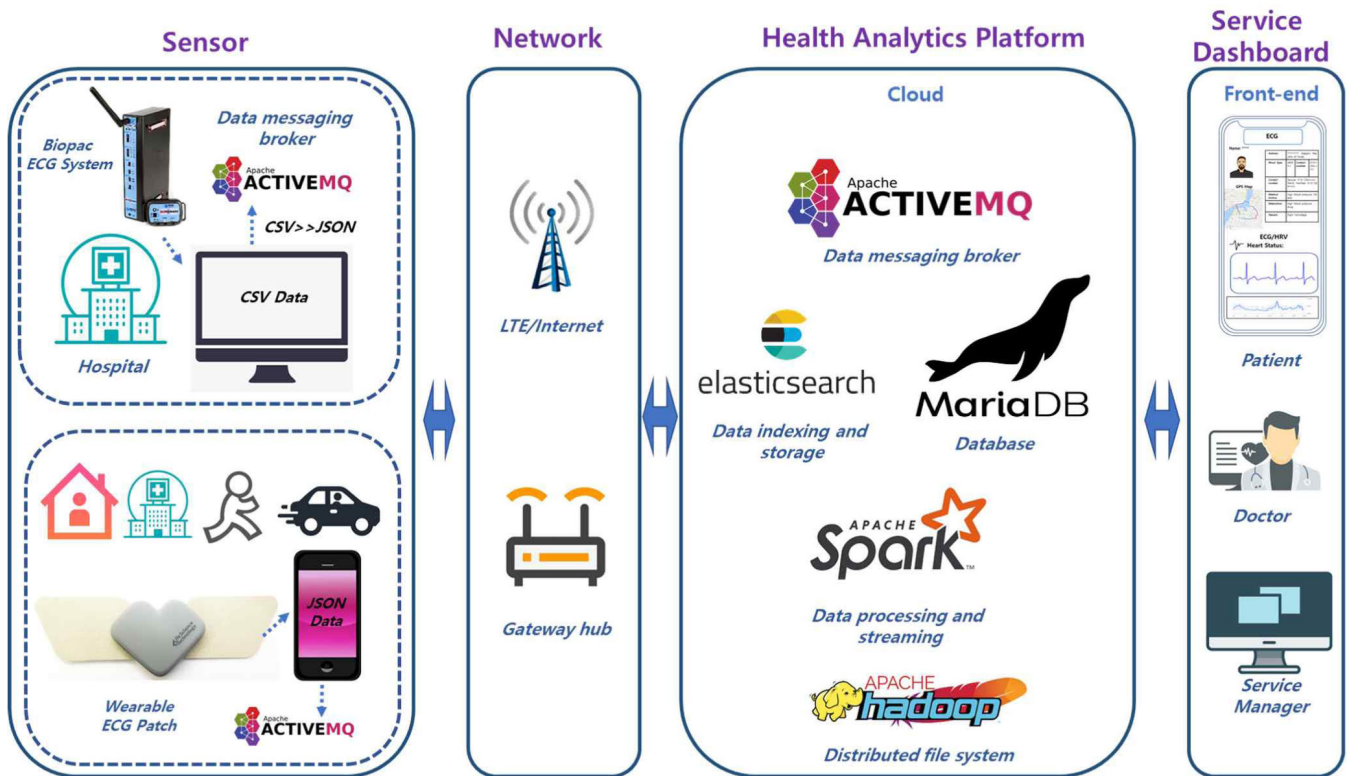


FIGURE 1. Overview of the Big-ECG system. ECG data acquisition system consists of the standard clinical ECG device and the wearable ECG patch. System connects the ECG patch with the phone API through BLE. System feed the ECG data to cloud server using Wi-Fi or LTE network. In Apache Hadoop based distributed file system, Elasticsearch indexes the data and acts as the NoSQL database, Spark performs live data processing, and MariaDB acts as the relational database (RDB) and provides query service for front-end service application. This ambulatory system is developed to identify the changes in cardiac features due to ischemic stroke or other illnesses and generate health advise and messages to assist the patients.

medical advice as feedback. This section describes the cyber-physical cardiac monitoring system, the sensor (physical) system, the cloud management system, and the front-end service dashboard.

1) SENSORS SYSTEM

A sensor is a device, module, machine, or subsystem whose purpose is to detect physical environment changes. The Healthcare system utilizes a wide range of biosensors to record various physiological signals to understand the health status. Our sensor system deals with ECG sensors: the wearable ECG patch for real-life and clinical applications and the standard ECG equipment for clinical application. The wearable ECG device is a self-powered ECG patch with Bluetooth low energy (BLE) communication with a computer or a phone. The standalone ECG equipment used in hospitals and healthcare centers stores the data in the local computer. We integrated both kinds of devices with our CPS system.

2) CLOUD DATA STORAGE AND PROCESSING

The physiological data, such as ECG can be utilized as big data, which is characterized by 5Vs in connection with Volume, Velocity, Variety, Value, and Veracity. Hospital-generated patient physiological data are of petabytes or zettabytes, which depict the volume. The velocity is stated

in terms of data sampling rate from the patients, and most of the clinical data are recorded with a higher sampling rate to ensure signal quality. Variety explains the diversified data sets, such as physiological data (ECG, EEG, EMG, etc.), and radiological images (MRI, CT), and veracity explains the data sets' reliability and availability. The recorded healthcare data are transformed into meaningful perceptions, such as disease prediction, health monitoring, disability assistance system, which describe the value in 5Vs.

The Cloud management system includes the data acquisition system, data processing, data storage, data serving. The management of volume, velocity, scalability, and fault-tolerance is the cloud platform's essential requirement. We utilized the Apache ActiveMQ for the role of data acquisition requirements. ActiveMQ is a message broker built on top of Java Messaging Service, capable of sending messages between applications [24]. A custom-made java-based sensor API acts as the publisher of the data. ActiveMQ is responsible for the transport of the data sent by the API into the cloud. ActiveMQ reduces message loss utilizing its fault-tolerance functionality. Elasticsearch is a NoSQL database, which acts as a distributed storage, search, and analytics engine with an HTTP web interface and JavaScript Object Notation (JSON) documents [25]. It provides powerful APIs to index data in a format of a dynamic number of key-value pairs. Logstash is a freely available data transformation engine that consumes

data from many sources, converts it into JSON format, and feeds it to the Elasticsearch database. The Hadoop ecosystem contains a Hadoop distributed file system (HDFS), MapReduce, Spark Streaming, and many other analytical components for solving Big Data problems, and they have become mature and enterprise-ready. The HDFS is designed for reliable storage, managing huge files, and streaming those data sets to front-end applications. Apache Spark Streaming is a scalable, fault-tolerant data processing for live data streams. Spark performs in-memory big data processing with low-latency; it is considered the best solution for live sensor data processing. MariaDB is a community-developed MySQL relational data management system. In this RDB, mostly structured data is processed and utilized in conjunction with Hadoop. It stores processed data and makes it available to use from the front-end dashboard with no latency in response.

3) MEDICAL ONTOLOGY

An ontology is a formal, explicit specification of a shared conceptualization [26]. As the ontology concept is gained attraction in the biomedical domain for knowledge interpretation and semantic interoperability, we developed the medical ontology with assistance from ontology experts and domain experts (researchers and doctors). The Protégé-OWL v.4.2 ontology editor [27], which supports the OWL (Ontology Web Language), was utilized to implement the ontology concepts. Stored ECG data are tagged in semantic annotations with predefined metadata, the set of ontological concepts. Semantically annotated data was stored in a resource description framework (RDF) database as RDF triples. The RDF database acts as the ontology engine and facilitates the storage and recovery of RDF triples through semantic queries. The Front-end knowledgebase system generates automated recommendation from the back-end ontology model for stroke prognostics, correlation with physiology, level of stroke recovery, post-stroke therapy, etc. In addition, the patient monitoring system subscriber or therapist can ask additional queries through user interface. The medical ontology shares the disease information and the correlation between the illnesses and physiological outcomes. Therefore, Ontology-based stroke prognostics and risk will appear in the client apps or dashboard.

4) HEALTH ADVISOR SERVICE DASHBOARD

The health advisor service dashboard consists of the client application, the clinical dashboard, and the service executive dashboard. The health advisor is the healthcare service layer including client profile nodes, such as personal information, wearable sensor identification, historical health records; resource nodes, such as hospital and emergency service information, hospital service availability, doctor information, medicine, hospital logistics; decision nodes, such as medical ontology, knowledgebase, disease ontology. The real-time or near-real-time biosignal monitoring, live data streaming and processing in a cloud platform, the real-time health status feedback to the client, and the hospital dashboard make

a way to automate the health advisor system. Our system enables the medical experts, such as doctors, to verify the automatically generated recommendation and add his expert clinical recommendation through the clinical dashboard. The health advisor service, shown in client apps or dashboards, consists of predicted diseases and severity and advice based on knowledgebase along with a doctor's prescription. The health advisor also provides a message service to emergency service control rooms and relatives about cardiac health, helping them assist their patients and move to the hospital for additional diagnosis and treatment.

B. DATAFLOW

The ECG data flows from the wearable sensor to front-end visualization and undergoes a series of data processing for the patients' real-time cardiac monitoring to detect ECG changes due to illness. Here we will describe the type of ECG data, travel route of data for the transformation of raw data to cardiac features, rule-based and machine learning-based data processing, data visualization in dashboards.

As demonstrated in Figure 2, the wearable ECG patch communicates with an API using the BLE network in a near-located android phone. The android application reads real-time ECG data from sensors, publishes in ActiveMQ topics, and feeds data in JavaScript Object Notation (JSON) format to the Transmission Control Protocol (TCP) server through Wi-Fi or Long-Term Evolution (LTE) network. ECG data is annotated with the corresponding device identity number (ID), patient ID, gateway ID, and timestamp to make data traceable. Besides, the standalone ECG equipment stores the data in comma-separated value (CSV) format. In this system, ECG CSV data is converted to JSON format and send to the ActiveMQ queue. JSON files were published by data acquisition API in ActiveMQ feed to the server. If the big data server exists on the same computer, Logstash can transform those CSV files into JSON format, suitable for Elasticsearch indexing and management.

On the cloud side, Elasticsearch receives raw data and performs indexing according to ECG data configuration protocol. The Spark streaming service accomplishes the processing of the live ECG. Data processing methods include context prediction, feature extraction, feature ranking, machine learning, and knowledgebase. The context predictor annotated the data with the client's situation (resting, active), activity information (walking, driving, or sleeping). The feature extraction module extracts all relevant cardiovascular features. The rule-based feature extender annotated cardiac features according to the predicted diseases, such as ischemic stroke. All the disease prediction rules come out of the disease prediction decision tree derived from early extensive clinical studies. The selected ECG features run through the machine learning model to train the model. The system keeps records of the clients' details, historical medical records, contact information, and health insurance data in their portfolios. All the processing data are stored in the MariaDB and made available for the front-end dashboard. The dashboard can communicate

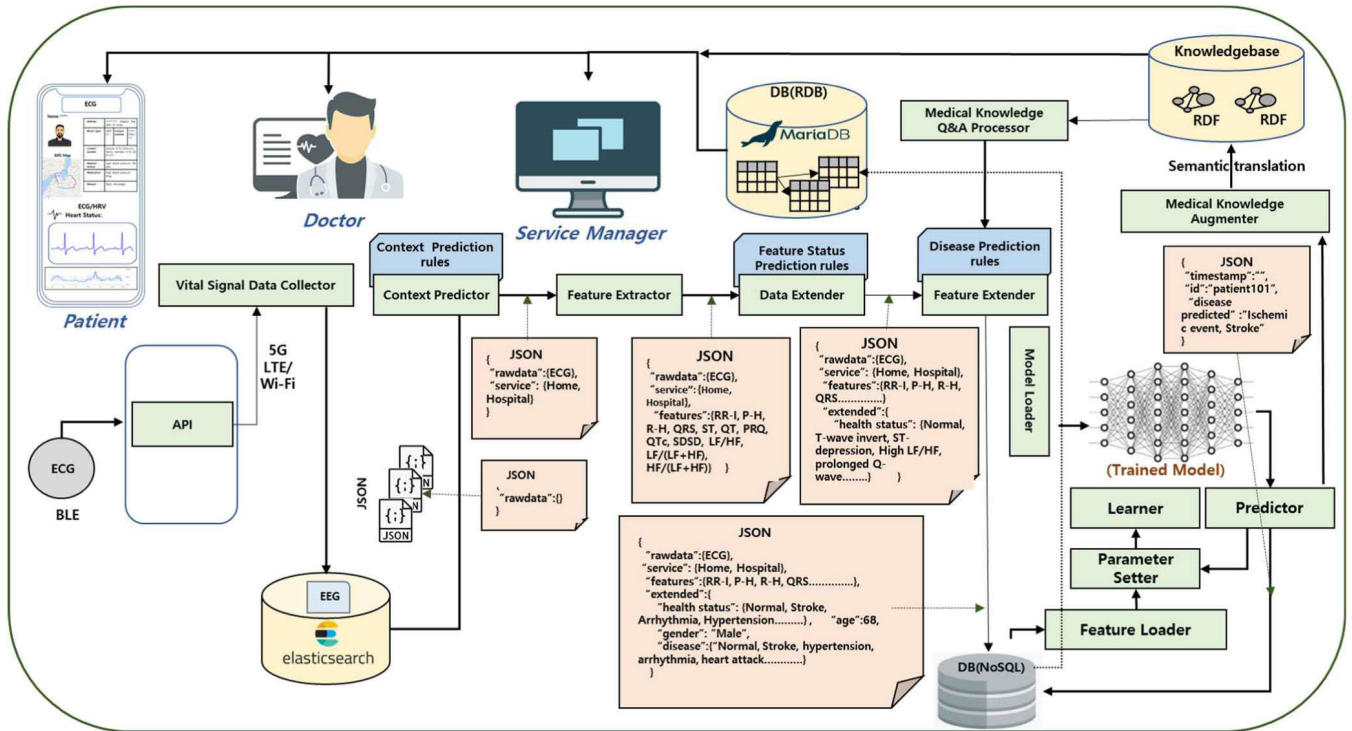


FIGURE 2. The dataflow of the Big-ECG system. System feed the ECG data to cloud server through Wi-Fi or LTE network using ActiveMQ queue. In cloud server, Elasticsearch indexes and stores the data, Spark performs live data processing, such as context prediction, feature extraction, rule-based feature extension, and machine learning based prediction. Context prediction node identifies the scenario of data. System extracts cardiovascular features related to Stroke, followed by the feature extension based on disease prediction rules. Cardiac features with disease prediction feed to machine learning model for training the model to build a disease prediction engine. RDB stores the processed data and provide query service for front-end service application. Disease ontology will assist to understand the correlation of physiology, diseases, and possible cause of diseases. Doctors can recommend expert suggestion through clinical dashboard.

with cloud applications in two ways; one is a REST API to query the dashboard data, and another one is a WebSocket, which streams direct messages to the dashboard. The Big-ECG system displays the cardiovascular health status and various key ECG and HRV features, such as RR interval, ST, QT, LF/HF ratio in the dashboard through WebSocket, and signal trends using an HTTP request. Medical Ontology and health advisors serve possible health advice to guide the patient and the healthcare service providers.

IV. EXPERIMENTAL METHODOLOGY

To understand the stroke-impaired cardiac activity, we measured the single-channel ECG of the stroke patients and the healthy adults in the resting state. We processed and extracted the ECG fiducial features, time-domain, and frequency-domain features of ECG-derived HRV. We investigated the cardiac features through statistical analysis and hypothesis tests to identify the significant important ECG features associated with ischemic stroke. We also utilized machine learning algorithms to automate the classification of stroke group and control group. As a pilot system, we set up a cyber-physical pilot system for stroke prognostics and rehabilitation management parallel to regular operations in two medical centers. The pilot system includes data acquisition using a wireless ECG patch, data transfer to the cloud, the wireless network, and data storage and indexing in the cloud platform.

A. DEMOGRAPHICS OF THE PARTICIPANTS

The participants of this experiment are ischemic stroke patients and healthy adults. The stroke group consisted of 45 ischemic stroke patients (Age: 70.8 ± 4.6 years old, 68% men), and the control group composed of 40 healthy adults (Age: 75.4 ± 2.3 years old, 38% males). Although no changes were observed for age and gender in ECG autonomic response [28], both participants in the stroke and control volunteers belong to the same age group to reduce age-related ECG fiducial feature variations. The stroke group included patients undergoing post-stroke rehabilitation at Chungnam National University Hospital and Konyang University Medical campus in Daejeon, South Korea. CT or MRI confirmed clinical diagnostics of the patients’ ischemic stroke. The control group consisted of healthy older adults with no underlying known heart disease and records of ischemic events. The Institutional review Board of the Korea Research Institute of Standards and Science, Daejeon, South Korea, and Konyang University, Daejeon, South Korea, approved this study conducted under the guidelines of the Declaration of Helsinki (KRISS-IRB-2016-05-19).

B. ECG DATA ACQUISITION

We recorded ECG data in two different sensor systems; one is the Biopac wireless ECG sensor (Biopac Systems Inc., Santa Barbara, CA, USA), and another is a wearable ECG patch

(Life science Technology Inc., South Korea). We acquired a single-channel ECG dataset in Chungnam National University Hospital using the Biopac MP160 system with AcqKnowledge version 5.0. A wireless Biopac BioNomadix respiration (RSP) and ECG amplifier (RSPEC-4.3) recorded the cardiovascular activity using 3 x 30-cm Electro Lead (BN-EL30-LEAD3) by applying bipolar EL 503 pre-gelled disposable electrodes to the left and right chests of the participants, as shown in Figure 3(c). We recorded a single-channel ECG dataset in Konyang University Physiotherapy Center using an ECG patch and feed it to the cloud database. We used the low-alcohol swab to clean the participants' skin to reduce the impedance. As described in Figure 3(a), we only consider ECG data gathered on the lead position V5. For the stroke population, we recorded the ECG data within three months after diagnosing Ischemic Stroke. We recommended participants avoid drinks, such as coffee or alcohol, before the recording. While measuring ECG data, we instructed the patient to keep awake, close the eyes, sit down and keep rest. Following sitting on the chair, the recording of the data was delayed for 3 minutes, allowing the participant's vital signs to calm down to a steady-state. As demonstrated in Figure 3(b), we recorded the electrocardiogram for at least 5 minutes during awake and rest. We maintained the room temperature at 24° C and the relative humidity at 40%.

C. DATA TRANSFORMATION AND STORAGE

We fed the ECG patch data directly to the Big-ECG server through an android application using ActiveMQ. Besides, ECG data of the Biopac wireless is stored ECG data in CSV file format in the connected local computer. To feed this data to a remote server using the ActiveMQ protocol, a data conversion API transformed the CSV data to JSON data and sent it to the ActiveMQ queue. On the remote server, Elasticsearch receives raw data and makes indexing according to ECG data configuration protocol. As displayed in Figure 3(d), the data server is equipped with a Dell PowerEdge T640 tower server (Intel Xeon Silver 4210R 2.4GHz 10C Processor, RAM:32GB).

D. PRE-PROCESSING

All electrocardiogram (ECG) streams were sampled down at 200 Hz to match the optimized sampling rate of the QRS detection algorithms. All premature, missing, or ectopic beats are filtered out using Pan-Tompkins QRS detection algorithm [29].

E. FEATURE EXTRACTION

ECG Feature extraction consists of the fiducial features and the heart rate variability features as described in Figure 3(e). ECG fiducial components extracted through the onset, offset, and peak of each wave of the standard P-QRS-T wave profile. We analyzed ECG-derived HRV signals in time-domain and frequency-domain. The frequency-domain HRV features are spectral power extracted in various frequency bands, and the

time-domain HRV features included the various statistical components.

1) ECG FIDUCIAL FEATURES

We extracted the fiducial features from the ECG waveform. The cycle-by-cycle time and voltage measurements of Q and S wave events and QRS events are extracts for various points and intervals between waveforms in the ECG signals cycle. RR Interval demonstrates the time between successive R peaks in the ECG waveform calculated in seconds. Heart rate, expressed in beat per minute (BPM), is calculated using the RR time interval. QRS defines the duration between the start of the Q -wave and the end of the S-wave. QT describes the period between the beginning of the Q wave and the end of the T-wave measured in seconds. Corrected QT interval (QTc) is the QT duration adjusted with the RR interval. ST describes the time between the S wave and the end of the T wave calculated in seconds. PRQ interval means the period between the beginning of the P-wave and the Q-wave measured in seconds. P-height (P-H) narrates the height of the P-wave peak in a cycle measured in mV. Similarly, R-height (R-H) expresses the R-wave amplitude in an ECG cycle recorded in mV.

2) TIME-DOMAIN HEART RATE VARIABILITY

HRV is a measure of the physiological rhythm between successive beats. The change in heart rate is detected in the RR interval of the ECG waveform. The RR interval is a representative function of heart rate (HR) and HRV. The RR interval is extracted from the ECG signal using a QRS detector. A modified Pan-Tompkins method is used to normalize the ECG data to 1, rather than using raw ECG data, where the peak value of the highest R-wave is considered 1 [29]. R wave threshold is demonstrated in normalized range (-1, 1): positive for positive R wave peaks and negative for inverted R peaks. A continuous time-domain representation of the RR Intervals is obtained through re-sampling R-R intervals to a constant sampling rate using the Cubic-spine interpolation. The features obtained in the time domain analysis were generally the standard deviation of the adjacent R-wave interval (SDNN), the RMS of the successive difference of the RR interval (RMSSD), and the standard deviation of the consecutive difference of the R-wave interval (SDSD) and the ratio of the number of pairs of normal-to-normal R-wave (NN) intervals greater than 50 ms to the total NN interval (pNN50). Respiratory Sinus Arrhythmia (RSA) is an index for the respiratory cycle, defined as maximum rate minus the minimum rate, expressed in milliseconds.

3) FREQUENCY-DOMAIN HEART RATE VARIABILITY

The power spectral density (PSD), power in various frequency bands are extracted from the RR intervals using the Welch approximation method, the average of signal time-sliced portions. The Hamming window is used for Fast Fourier transformation (FFT) to construct PSD. VLF, the very-low-frequency band power, describes the average

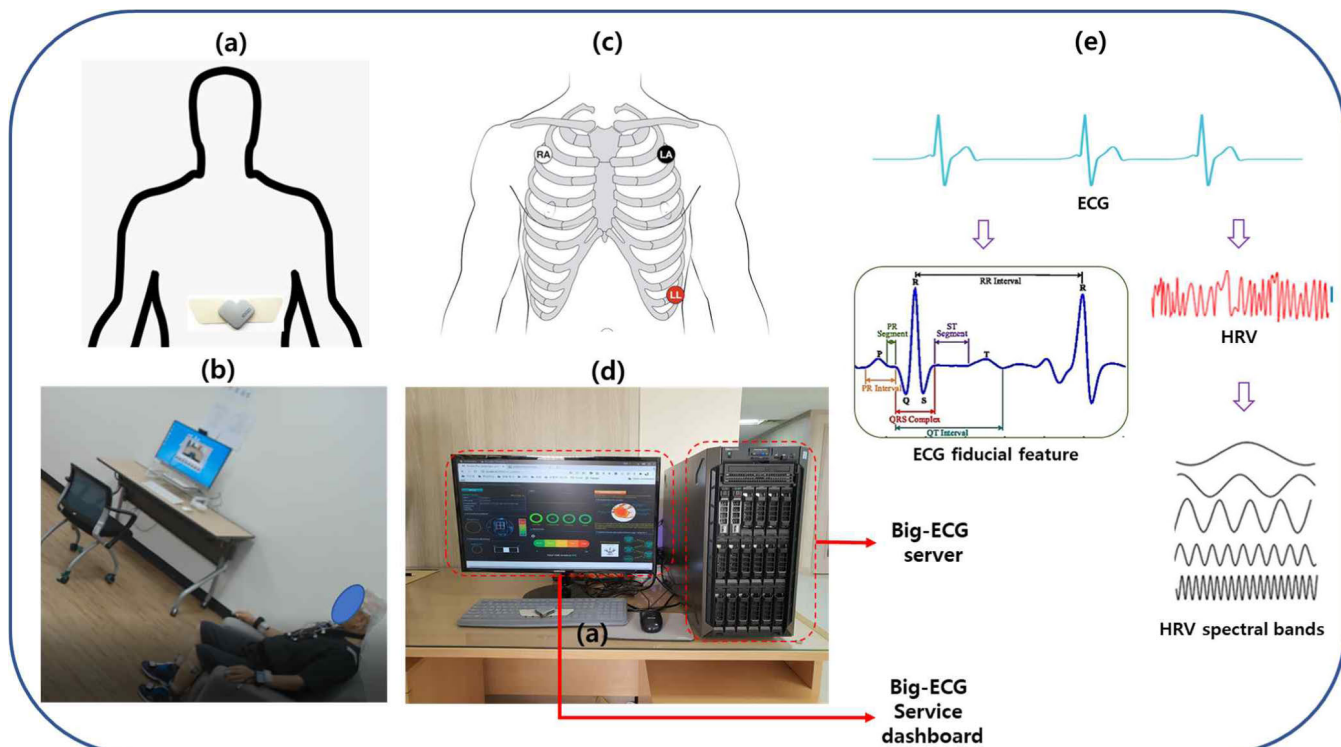


FIGURE 3. Demonstration of data acquisition protocol, experimental scenario, and signal processing. (a) ECG patch in V5 position for both clinical and real-life applications; (b) Experimental scenario; (c) Biopac system ECG electrode in V5 position for clinical application; (d) Big-ECG platform and service dashboard; (e) Standard ECG signal and fiducial points description.

spectral power measured in the range of 0.00-0.04Hz, having a unit of $\text{second}^2/\text{Hz}$. LF, the low-frequency band power, is the average spectral power measured in the range of 0.04-0.15Hz. HF, the high-frequency band power, indicates the average spectral power measured in the range of 0.15-0.40Hz. VHF, the very-high-frequency band power, demonstrates the average spectral power measured in the range of 0.40-3.00Hz. $\text{LF}/(\text{LF} + \text{HF})$ narrates the low-frequency ratio, and $\text{HF}/(\text{LF} + \text{HF})$ mentions the high-frequency ratio. LF/HF describes the ratio of low-frequency power and high-frequency power.

F. FEATURE SELECTION

Feature selection plays a vital role in high-dimensional biomedical data analysis. Classification performance largely depends on the relevance of features, and irrelevant or redundant data affects the computational power and time. Feature selection consists of screening, ranking, and selecting features. Screening removes feature variables, which do not provide useful information for prediction. Feature selection ranks the features based on the prediction accuracy of the individual variable. The chi-square test measures the importance value of the predictor. We evaluated the feature importance as $(1-p)$, where p is the chi-square test outcome. We selected ECG features with feature importance greater than 0.95 for training the machine learning algorithm.

G. CLASSIFICATION

Supervised machine learning techniques are an efficient tool for classification and discovering patterns in a dataset. In previous studies, machine learning was successfully utilized to classify the physiological [21], [30] and behavioral [20], [31] data of the stroke dataset and the control dataset. Machine learning and deep learning techniques are also utilized to classify the fatigue indices [32] and sleep apnea [33] using the multimodal physiological signal. Decision tree-based machine learning algorithms, such as QUEST, CART, C5.0, CHAID, Random Trees, and biologically inspired neural networks algorithm, have been implemented to categorize cardiac stroke features. ECG HRV features were extracted for every epoch of 30s and, fiducial features were generated for each QRS cycle. We also filtered the premature, missing, or ectopic beats and corresponding epoch measurements. ECG dataset consists of 521 sets of HRV pre-processed features and 5658 sets of pre-processed fiducial features in total; 6 sets of HRV features and 66 sets of fiducial features on average for each subject. ECG HRV dataset has 365 sets of HRV features and each set consists of 5 time-domain HRV features, such as RMSSD, SDDSD, pNN50, RSA, and 9 frequency-domain HRV features, such as VLF, LF, HF, VHF, LF/HF , $\text{LF}/(\text{VLF} + \text{LF} + \text{HF})$, $\text{HF}/(\text{VLF} + \text{LF} + \text{HF})$, $\text{LF}/(\text{LF} + \text{HF})$, and $\text{HF}/(\text{LF} + \text{HF})$, extracted for each sample. Furthermore, ECG fiducial dataset has 3961 sets of fiducial features and each set consists of 8 fiducial or

TABLE 1. Model hyperparameters of machine learning models (CHAID, QUEST, CART, C5.0, Random trees, Neural networks).

CHAID	QUEST
Decision Tree developing algorithm: CHAID Tree depth limit (Maximum): 5 Maximum surrogates: 5 Minimum records in parent branch (%): 2 Minimum records in child branch (%): 1 Number of Ensemble component models for boosting and/or bagging: 10 Overfit prevention set (%): 30 Significance level for splitting and merging: 0.05 Method for adjusting significance values: Boferroni method Categorical targets prediction: Pearson Chi-square method Minimum change in expected cell frequencies: 0.001 Maximum iterations for convergence: 100 Minimum change in impurity: 0.0001	Decision Tree developing algorithm: QUEST Tree depth limit (Maximum): 5 Maximum surrogates: 5 Minimum records in parent branch (%): 2 Minimum records in child branch (%): 1 Number of Ensemble component models for boosting and/or bagging: 10 Overfit prevention set (%): 30 Significance level for splitting and merging: 0.05 Method for adjusting significance values: Boferroni method Categorical targets prediction: Pearson Chi-square method Minimum change in expected cell frequencies: 0.001 Maximum iterations for convergence: 100 Minimum change in impurity: 0.0001
CART	C5.0
Decision Tree developing algorithm: CART Tree depth limit (Maximum): 5 Maximum surrogates: 5 Minimum records in parent branch (%): 2 Minimum records in child branch (%): 1 Number of Ensemble component models for boosting and/or bagging: 10 Overfit prevention set (%): 30 Significance level for splitting and merging: 0.05 Method for adjusting significance values: Boferroni method Categorical targets prediction: Pearson Chi-square method Minimum change in expected cell frequencies: 0.001 Maximum iterations for convergence: 100 Minimum change in impurity: 0.0001	Decision Tree developing algorithm: C5.0 Tree depth: 22 Expected noise (%): 0 Cross-validation: yes (10-folds) Favor: Accuracy Output type: Decision tree Use boosting: True Number of trials: 10
Random Trees	Neural Networks
Decision Tree developing algorithm: Random Trees Tree depth limit (Maximum): 10 Minimum child node size: 5 Maximum percentage of missing value: 70 Maximum number of field categories: 49 Minimum field variation: 0.05 Number of bins: 10 Maximum number of nodes: 10000	NN model: Multilayer perceptron (MLP) Stopping rules: Use maximum training time (per component model): 15 minutes Combining rule for continuous targets: Mean Number of component models for boosting and/or bagging: 10 Overfit prevention set (%): 30

profile features, such as RRI, R-H, P-H, QRS, PRQ, QT, QTc, and ST extracted for each sample. We partitioned the ECG dataset into the training and testing data. The training dataset comprises 70% of feature data, and the test dataset occupied 30% of the entire feature dataset. Training data size is 365 sets of HRV features and 3961 sets of fiducial features. Besides, the testing data size was 156 sets of HRV features and 1697 sets of fiducial features. We tuned the hyper-parameters of models using cross-validation to find the best-performing model. We performed non-exhaustive k-fold (k = 10) cross-validation using the training dataset to get rid of overfitting [34]. Each model was trained and cross-validated to find out the set of hyper-parameters with the highest accuracy of the model. As the most accurate model was developed, we test the model using the test dataset. The optimized hyper-parameters of each model were presented in Table 1.

1) CHAID MODELS

The chi-squared automatic interaction detector (CHAID) method is a decision tree formed by successively dividing a subset into two or more child nodes, starting with the whole

data set [35]. The best partition across all nodes comes out by merging the predictors’ pairs until no significant difference is observed within the target’s pair. As a decision tree model, CHAID model output is visual and easy to interpret in the clinical decision support system. c5.0 model The C5.0 model is a supervised data mining algorithm used to build decision trees from data sets. It creates a decision tree using a divide-and-conquer method. The C5.0 decision tree algorithm uses a gain ratio as the basis for division. The model builds the decision tree, followed by the cleanup procedure and the tree size reduction to minimize the tree’s estimation error rate [36]. This algorithm is widely utilized in biomedical data mining applications.

2) QUEST MODEL

QUEST (Quick, Unbiased, Efficient) is a binary-split statistical tree-growth method [37]. QUEST handles linear splits using Fisher’s Linear discriminant analysis. If no missing values in the data, it grows a tree with univariate splits. It is robust to handle categorical predictors with many categories.

3) NEURAL NETWORK MODEL

The neural network is a biologically-inspired data mining algorithms that predict a target according to a growing multi-layered intricate pattern. We used the multilayer perceptron (MLP) neural network in this study [38]. This model includes an input layer with multiple input nodes, a neural network with hidden layers, and an output layer. This model is capable of learning by own, fault-tolerance and storing the data in entire network, capable of working on real-time applications.

4) CLASSIFICATION AND REGRESSION TREES MODEL

Classification & Regression Tree (CART) is a recursive segmentation method suitable for regression and classification by selecting partitions at each node. Each child node created by the separation is more homogeneous than the parent node [39].

5) RANDOM TREES MODEL

The random trees model is a robust supervised Classifier for accurate predictive models in classification or regression problems. Random Trees is an ensemble learning algorithm consisting of tree nodes representing decision rules to understand any tree's prediction and generate multiple classifications and regression trees [17].

H. DATA ANALYSIS

We explored the cardiac features that characterize ECG changes due to ischemic stroke using statistical and machine learning data analysis. We performed the statistical analysis to identify the relationship between ECG-derived variables. We explored the descriptive statistics analysis to explore the statistical distribution of the data and independent-samples t-test to evaluate whatever the associated groups' means are statistically significant. We performed the Statistical analysis using SPSS 24 package (IBM, Armonk, NY, USA). Machine learning techniques are practical in assessing the most accurate predictions possible. We used the feature selection to rank ECG features based on the target prediction performance. For feature selection, Pearson's chi-square test evaluated the prediction importance of the component. The supervised machine learning algorithms utilized the high-ranking training feature datasets to build a classification model, which later tested the dataset. We used the IBM SPSS Modeler 18 package (IBM, Armonk, NY, USA) to utilize machine learning techniques in our ECG data.

V. RESULTS

We developed a real-time or near-real-time ECG-based health monitoring and disease prediction platform. The core modules are a wearable ECG patch for cardiac signal acquisition, a big data platform for real-time data storage and processing, and the health advisor dashboard for post-stroke management service. We investigated the association of the electrocardiographic features with post-stroke ECG in two methods.

(1) Statistical analysis included descriptive statistics and the hypothesis test. Descriptive statistics provide statistical distribution measures, such as mean, variance, standard deviation. In descriptive statistics, a boxplot graphically portrays the spread of the dataset with their quartiles. The independent sample t-test is a hypothesis test to determine whether associated population means are statistically different. We performed Levene's test to measure the equality of the variance and the t-test to check the means' equality. (2) The machine learning technique is a data analysis method, which builds analytical models to learn from data, identify patterns and make decisions through experience. In the following subsections, we will explore the results of the descriptive statistics and the hypothesis tests of important ECG fiducial and the heart rate variability features.

A. ASSOCIATION BETWEEN ECG FIDUCIAL FEATURES AND STROKE

As displayed in Figure 4, RR interval, P-height, QRS, QT, QTc, and ST intervals are the most significant stroke-predictive ECG features. We investigated whether post-stroke ECG changes are associated with the ECG fiducial features and whether these can be detected using a single-channel heart signal recording. As shown in Table 2, the RR interval was -0.025 s shorter in the stroke group relative to the control group (95% CI, -0.031 to -0.018 s, $p = 0.0001$). Mean R-H was 0.085 mV higher in the stroke group relative to the control group (95% CI, 0.055 - 0.115 mV, $p = 0.0001$). Mean P-H was 0.011 mV higher in the stroke group relative to the control group (95% CI, 0.010 - 0.012 mV, $p = 0.0001$). The mean QRS of the stroke group was 0.004 s longer (95% CI, 0.002 - 0.005 s, $p = 0.0001$) than the mean QRS for the control group. The mean PRQ of the stroke group was -0.006 s shorter (95% CI, -0.007 to -0.004 s, $p = 0.0001$) than the mean QRS for the control group. The mean QT of the stroke dataset was -0.018 s shorter (95% CI, -0.021 to -0.015 s, $p = 0.0001$) relative to the control group's mean QT. The mean QTc of the stroke patients was -0.011 s shorter (95% CI, -0.014 to -0.008 s, $p = 0.0001$) relative to the control patients' mean QT. The mean ST of the stroke group was -0.023 s shorter (95% CI, -0.027 to -0.020 s, $p = 0.0001$) than the control group's mean ST. Although all ECG fiducial variables' mean values were significantly different, the variance values of R-height, PRQ interval showed discrepancies.

B. ASSOCIATION BETWEEN FREQUENCY-DOMAIN HEART RATE VARIABILITY AND STROKE

As demonstrated in Figure 5, LF ratio, HF ratio, LF/HF showed significant associations with post-stroke cardiovascular activity. We conducted the statistical investigation to evaluate the association of frequency-domain features of HRV with the stroke group relative to the control group. As listed in Table 3, we measured spectral power in the LF, HF, VLF, VHF bands and extracted the spectral ratios, such as LF/(LF + HF), LF/(VLF + LF + HF), HF/(LF + HF), HF/(VLF + LF + HF), and LF/HF as the standard

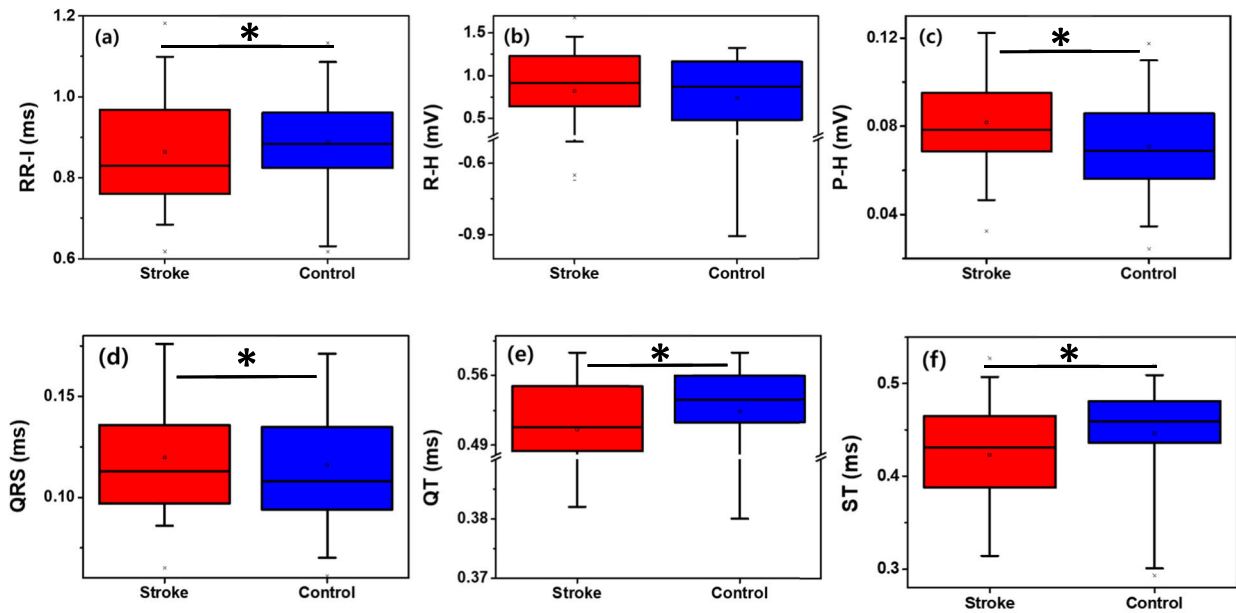


FIGURE 4. Statistical distribution of ECG fiducial features. Median and interquartile range of (a) RR interval; (b) R-height; (c) P-height; (d) QRS interval; (e) QT interval; (f) ST interval among the stroke group and the healthy control group. *($p < 0.05$) indicates significant difference.

TABLE 2. Results of statistical analysis of the ECG fiducial features of the stroke and control group. *($p < 0.05$) indicates significant difference.

ECG Fiducial Features	Mean Value		Mean difference	95% Confidence Interval		p-value (Levene's test for Equality of Variances)	p-value (t-test for Equality of Means)
	Stroke Group	Control Group		Lower	Upper		
RRI, s	0.86	0.89	-0.025	-0.031	-0.018	0.0001*	0.0001*
R-H, mV	0.82	0.74	0.085	0.055	0.115	0.167	0.0001*
P-H, mV	0.08	0.07	0.011	0.010	0.012	0.034*	0.0001*
QRS, s	0.12	0.12	0.004	0.002	0.005	0.017*	0.0001*
PRQ, s	0.17	0.18	-0.006	-0.007	-0.004	0.911	0.0001*
QT, s	0.51	0.52	-0.018	-0.021	-0.015	0.0001*	0.0001*
QTc, s	0.55	0.56	-0.011	-0.014	-0.008	0.0001*	0.0001*
ST, s	0.42	0.45	-0.023	-0.027	-0.020	0.0001*	0.0001*

TABLE 3. Results of statistical analysis of the HRV frequency-domain features of the stroke and control group. *($p < 0.05$) indicates significant difference.

HRV frequency-domain Feature	Mean Value		Mean difference	95% Confidence Interval		p-value (Levene's Test for Equality of Variances)	p-value (t-test for Equality of Means)
	Stroke Group	Control Group		Lower	Upper		
VLF, s ²	1.29	0.46	0.83	-1.42846	3.09032	0.140	0.470
LF, s ²	4.77	1.69	3.08	-5.27647	11.43048	0.140	0.470
HF, s ²	11.38	3.98	7.40	-12.51808	27.31310	0.137	0.466
VHF, s ²	31.24	14.03	17.22	-35.22255	69.65487	0.182	0.519
LF/HF	0.424	0.425	-0.001	-0.00141	-0.00001	0.012*	0.046*
LF/(VLF+LF+HF)	0.2757	0.2760	-0.0003	-0.00059	-0.00001	0.012*	0.044*
HF/(VLF+LF+HF)	0.6495	0.6491	0.0004	0.00001	0.00079	0.013*	0.045*
LF/(LF+HF)	0.2980	0.2983	-0.0003	-0.00070	-0.00001	0.012*	0.045*
HF/(LF+HF)	0.7020	0.7016	0.0004	0.00001	0.00070	0.012*	0.045*

HRV measures. Power in HF, LF, and VLF are not statistically significant ($p > 0.05$). The mean LF/(LF + HF) was -0.0003 smaller (95% CI, -0.0007 to -0.00001 , $p = 0.045$) in the stroke group relative to the control group. The mean HF/(LF + HF) was 0.0004 greater (95% CI, 0.00001 to

0.00070 , $p = 0.045$) in the stroke group relative to the control group. The mean LF/(VLF + LF + HF) was -0.0003 smaller (95% CI, -0.00059 to -0.00001 , $p = 0.044$) in the stroke group relative to the control group. The mean HF/(VLF + LF + HF) was 0.0004 greater in the stroke group relative to

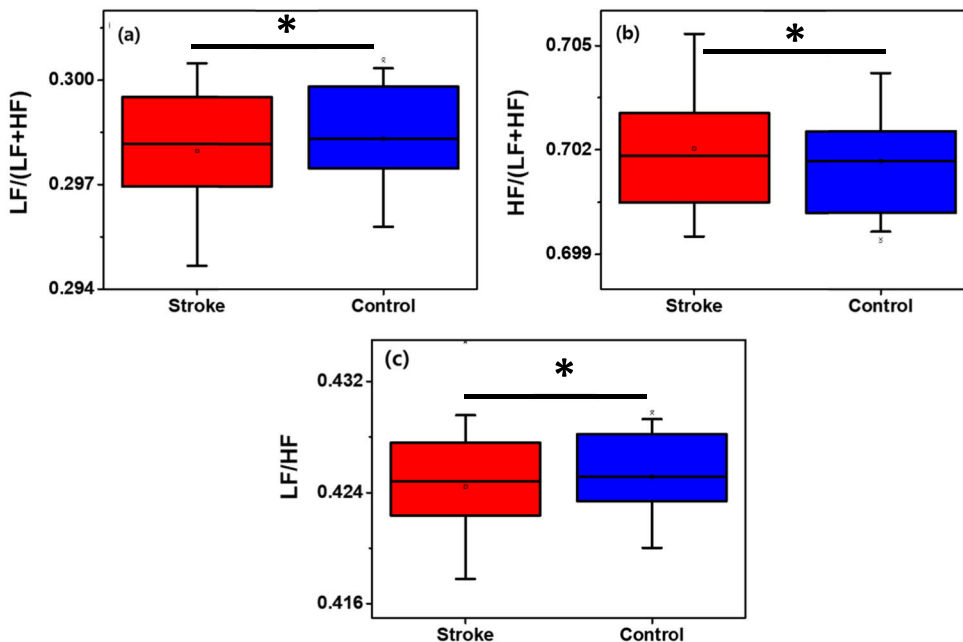


FIGURE 5. Statistical distribution of HRV frequency-domain features. Median and interquartile range of (a) LF/(LF + HF); (b) HF/(LF + HF); (c) LF/HF among the stroke group and the control group. *(p < 0.05) indicates significant difference.

TABLE 4. Results of statistical analysis of the HRV time-domain features of the stroke and control group. *(p < 0.05) indicates significant difference.

HRV time-domain Feature	Mean Value		Mean difference	95% Confidence Interval		p-value (Levene's Test for Equality of Variances)	p-value (t-test for Equality of Means)
	Stroke Group	Control Group		Lower	Upper		
RMSSD, ms	13.34	16.62	-3.28	-4.81	-1.75	0.107	0.0001*
SDSD, ms	12.92	16.39	-3.47	-4.98	-1.95	0.032*	0.0001*
pNN50, %	1.16	2.41	-1.24	-2.17	-0.32	0.0001*	0.008*
RSA, s	-1.24	-0.75	-0.49	-0.75	-0.23	0.086	0.0001*

the control group (95% CI, 0.00001 to 0.00079, p = 0.045). The mean LF/HF was -0.001 smaller (95% CI, -0.00141 to -0.00001, p = 0.046) in the stroke group relative to the control group.

C. ASSOCIATION BETWEEN TIME-DOMAIN HEART RATE VARIABILITY AND STROKE

As displayed in Figure 6, RSA, RMSSD, and SDSD are significant predictive features associated with cardiovascular activities after ischemic stroke. We investigated the statistical measure to evaluate the association of time-domain features of HRV with the stroke group relative to the control group. As shown in Table 4, we measured RMSSD, SDSD, pNN50, and RSA as the standard HRV measures. Few ECG-derived time-domain variables of heart rate variability have expressed association with stroke patients. The mean RMSSD of the stroke group was -3.28 ms shorter (95% CI, -4.81 to -1.75 s, p = 0.0001) than the control group’s mean RMSSD. The mean SDSD of the stroke group was -3.47 ms shorter (95% CI, -4.98 to -1.95 s, p = 0.0001) than the control group’s mean SDSD. The mean pNN50 of the stroke group was -1.24 % shorter (95% CI, - 2.17 to -0.32 %, p = 0.008)

than the control group’s mean pNN50. The mean RSA of the stroke group was -0.49 shorter (95% CI, -0.75 to 0.23, p = 0.0001) than the control group’s mean.

D. MACHINE LEARNING BASED POST-STROKE CARDIAC HEALTH PREDICTION

In the results of feature selection, seven features out of all ECG fiducial features and four features out of all ECG HRV features, ranked higher than 95% of the importance limit, are selected and feed to models. Receiver operating characteristic (ROC) analysis offers the most comprehensive description of prediction widely used in biomedical studies [40]. It shows all of the combinations of sensitivity and specificity that a machine learning model can deliver. AUC (area under the curve) is a performance indicator of the predictive model and defines the area under the ROC curve. The perfect score of the AUC is 1.0. The AUC less than 0.5 is not considered a useful classifier. Another alternative measure of AUC is the Gini coefficient, ranging between and 1, defined as two times (AUC-1). The confusion matrix or the error matrix delivers a complete representation of the predictions of true and false. We evaluated the standard performance measures, including

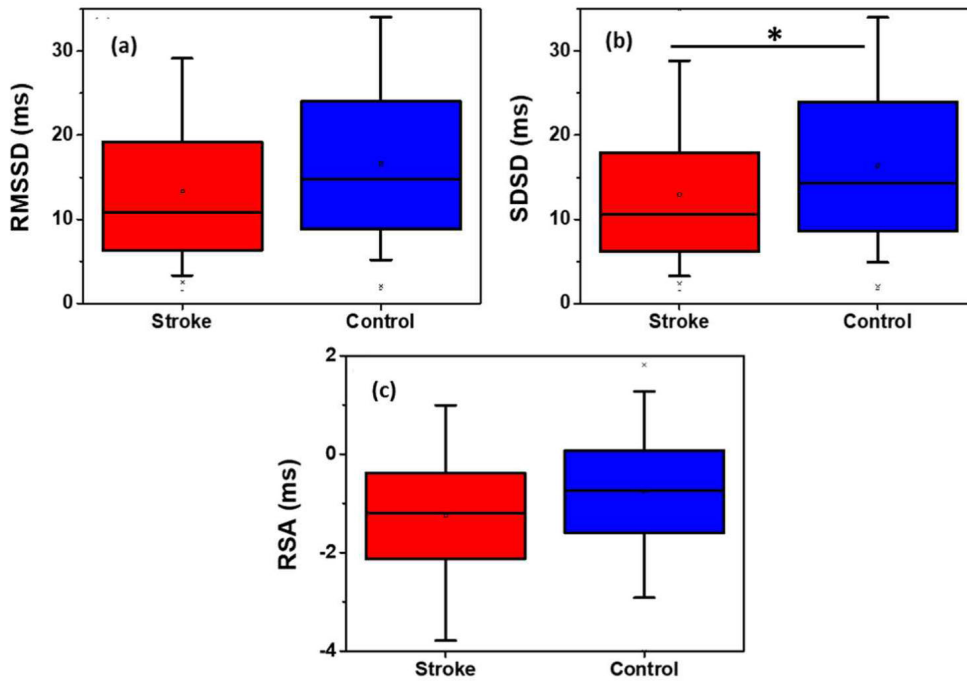


FIGURE 6. Statistical distribution of HRV time-domain features. Median and interquartile range of (a) RMSSD; (b) SDDSD; (c) RSA among the stroke population and the control population. *($p < 0.05$) indicates significant difference.

accuracy (ACC), sensitivity (true positive rate), specificity (true negative rate), precision (positive predictive rate), and negative predictive value from the confusion matrix. Accuracy was considered the most intuitive measure of performance to find the best model calculated as a percentage of the correct predictions across observations. Sensitivity is the true positive rate, defined as the correct positive predictive ratio of all actual observations. Specificity shows the true negative rate, characterized as the fraction of correct negative predictions to all actual observations. Model prediction outcome can be presented using the following standard equations:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Negative predictive value (NPV)} = \frac{TN}{TN + FN}$$

$$\text{Accuracy (ACC)} = \frac{TN + TP}{TN + TP + FN + FP}$$

where TP is a true positive, TN is a true negative, FP is a false positive, and FN is a false negative.

1) PREDICTION BASED ON ECG FIDUCIAL FEATURES

In Figure 7(a) and Figure 7(b), ROC curves demonstrate the classification models' performance curves using the training and the test datasets. All ECG fiducial features except QTc have shown feature importance greater than 0.95 for classification prediction in the feature selection. Table 5(a)

and Table 5(b) display all the classifiers' performance measurements for the training and test fiducial datasets. The Random Tree categorized the training dataset as the highest AUC (99.7%) and medium accuracy (ACC: 97.62%). The random trees model sorted the test datasets up to AUC (98.90%) and medium accuracy (ACC: 95.56%). As demonstrated in Figure 8, R-H, QRS, and PRQ have come out as the most crucial stroke classification features using the Random trees model. The C5.0 model categorized the training dataset with moderate AUC (99.5%) and accuracy (ACC: 98.20%). C5.0 classified test datasets by AUC (96.30%) and accuracy (ACC: 94.85%). RRI, R-H, and P-H have emerged as the most predictive stroke classification features using the C5.0 model. The CHAID model categorized the training dataset by AUC (99.40%) and accuracy (ACC: 97.59%). CHAID classified test datasets by AUC (96.90%) and accuracy (ACC: 94.27%). PH, ST, and QT have emerged as the most predictive stroke classification features using the CHAID model. The CART model categorized the training dataset by AUC (96.60%) and accuracy (ACC: 92.40%). CART categorized test datasets by AUC (94.70%) and accuracy (ACC: 91.23%). PH, ST, and QT have emerged as the most predictive stroke classification features using the CART model. The QUEST model categorized the training dataset by AUC (93.40%) and accuracy (ACC: 90.07%). QUEST classified the test dataset by AUC (92.10%) and accuracy (ACC: 89.18%). PH, ST, and QT have emerged as the most predictive features of stroke classification using neural network models. The neural network model categorized the training dataset with AUC (89.90%) and highest accuracy (ACC: 83.82%). The neural network

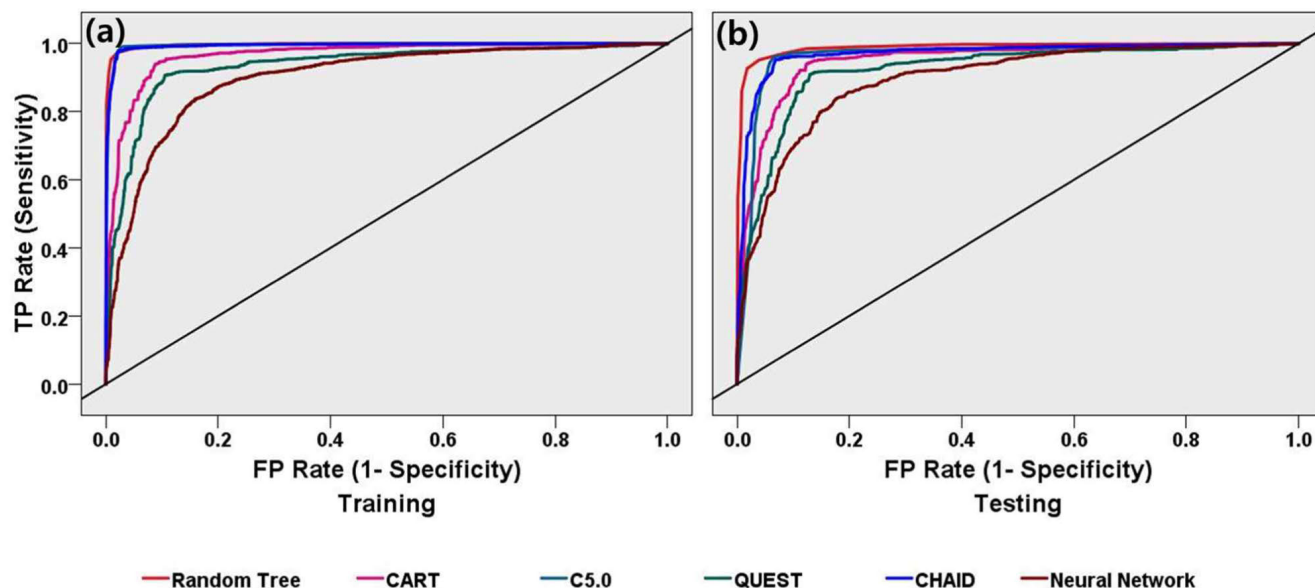


FIGURE 7. Receiver Operating Characteristic (ROC) curves for six different machine-learning models (Random Trees, CART, C5.0, QUEST, CHAID, Neural Network). Area under ROC curve (AUC) is an indicator of prediction accuracy. (a) ROC curve of the training dataset. Random Tree classified the training dataset with the highest AUC (99.70%) and moderate accuracy (ACC: 97.62%); (b) ROC curve of the testing dataset. Random Tree classified the testing dataset with the highest AUC (98.90%) and highest accuracy (ACC: 95.56%). Diagonal black line is the reference line.

TABLE 5. (a). Results of the classification performance of different Models using the ECG Fiducial training dataset. (b). Results of the Classification performance of different Models using the ECG Fiducial testing dataset.

(a)							
Model	Accuracy (ACC)	Sensitivity	Specificity	Precision	Negative Predictive Value	AUC	Gini
Random Trees	97.62	96.91	98.24	97.97	97.32	99.70	99.30
C5.0	98.20	98.97	97.31	97.72	98.77	99.50	99.00
CHAID	97.59	97.41	97.80	98.11	97.01	99.40	98.70
CART	92.40	95.06	89.30	91.20	93.94	96.60	93.10
QUEST	90.07	90.73	89.30	90.82	89.20	93.40	86.80
Neural Network	83.82	83.95	83.68	85.72	81.71	89.90	79.70

(b)							
Model	Accuracy (ACC)	Sensitivity	Specificity	Precision	Negative Predictive Value	AUC	Gini
Random Trees	95.56	95.19	96.04	96.98	93.74	98.90	97.70
C5.0	94.85	95.91	93.45	95.13	94.48	96.30	92.50
CHAID	94.27	95.09	93.18	94.89	93.43	96.90	93.80
CART	91.23	94.78	86.49	90.34	92.55	94.70	89.40
QUEST	89.18	90.89	86.90	90.24	87.74	92.10	84.30
Neural Network	83.17	83.91	82.17	86.30	79.23	89.40	78.80

categorized the test dataset by AUC (89.40%) and accuracy (ACC: 83.17%). ST, RR-I, and QT have emerged as the most predictive stroke classification features using neural network models. Overall, the random trees model shows the highest AUC (99.40%) and the highest Gini coefficient (98.90%).

2) PREDICTION BASED ON HEART RATE VARIABILITY

The ROC curves of machine learning models demonstrate the Stroke prediction performance using the HRV time-domain and frequency-domain features in Figure 9(a) and Figure 9(b). In the feature selection, RMSSD, RSA,

SDSD, and pNN50 have come out as the most predictive features (feature importance > 0.95) for stroke classification shown in Figure 10. Table 6(a) and Table 6(b) listed classifiers’ performance measurements using the training and the testing HRV feature dataset. The CART model classified the training dataset with the highest AUC (87%) and highest accuracy (ACC: 82%) and classified the testing dataset with the highest AUC (70%) and the best accuracy (ACC: 69%). CHAID categorized training datasets by AUC (77%) and accuracy (ACC: 70%) and test datasets by AUC (63%) and accuracy (ACC: 56%).

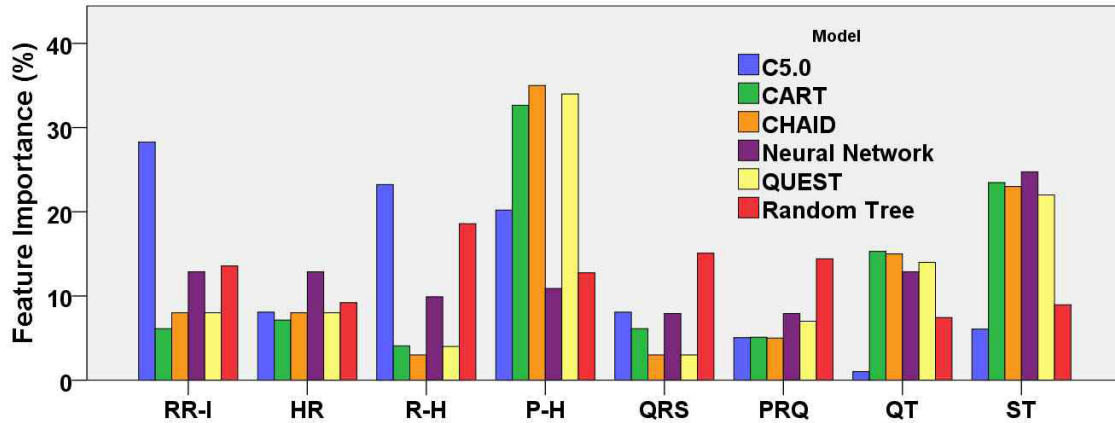


FIGURE 8. Feature importance of the ECG fiducial features in the feature ranking process of machine learning models to distinguish the stroke and control groups.

TABLE 6. (a). Results of the classification performance of machine learning models using the HRV training dataset. (b). Results of the Classification performance of Machine Learning Models using the HRV testing dataset.

(a)							
Model	Accuracy (ACC)	Sensitivity	Specificity	Precision	Negative Predictive Value	AUC	Gini
CART	0.82	0.83	0.80	0.84	0.79	0.87	0.74
CHAID	0.70	0.69	0.72	0.76	0.64	0.77	0.53
Random Trees	0.67	0.76	0.55	0.68	0.64	0.72	0.43
C5.0	0.66	0.68	0.62	0.70	0.61	0.69	0.38

(b)							
Model	Accuracy (ACC)	Sensitivity	Specificity	Precision	Negative Predictive Value	AUC	Gini
CART	0.69	0.68	0.71	0.78	0.59	0.70	0.40
CHAID	0.56	0.51	0.65	0.68	0.47	0.63	0.26
Random Trees	0.57	0.66	0.44	0.64	0.46	0.60	0.29
C5.0	0.58	0.57	0.60	0.68	0.48	0.61	0.22

VI. DISCUSSION

Our study aimed to investigate an ECG-based CPS feasibility and evaluate the cardiac biomarkers indicating activity changes due to ischemic stroke. Stroke shares severe health risk factors, and underlying heart diseases, such as heart failure, atrial fibrillation, or vascular heart disease increases stroke risk. Stroke impairs autonomic control and leads patients to cardiac complications [41], and post-stroke cardiac complications are the most deadly [42].

When an ischemic event, such as a hemorrhagic stroke occurs due to a rupture of blood cells, oxygen supply to the lesion area’s is disturbed and causes the brain cells to die. This damage to brain tissue affects the central nervous system [6] and the autonomic nervous system. ECG derived-HRV is one of the gateways for easy access to autonomic activity. Neurological disorders, such as acute ischemic stroke, change the ECG characteristics in various ways. ECG abnormalities may occur as complications, such as cardiac arrhythmia, such as ventricular tachycardia, ventricular tachycardia. Cardiac arrhythmias are responsible for hemodynamic

instability and responsible for unexpected sudden death after ischemic stroke. For example, atrial fibrillation, a kind of arrhythmias, can lead to subsequent brain and systemic thromboembolism [41].

According to this study, resting RR-I, QRS, QT, QTc, and ST are essential markers for classifying the stroke and healthy control groups. The primary and possibly deadliest ECG features associated with neurological illness are the ST-segment and T-wave, which reflect abnormal repolarization [47]. ST-segment depression [48], [49] is associated with ischemic stroke with underlying coronary heart disease. We observed a similar ST-pattern in our investigation. Prolongation of QTc is an independent predictor of reduced HRV and increases the threat of cardiac death in the stroke population [50], [51]. As our investigated stroke patients are in the recovery phase, a few stroke patients showed QT prolongation in this study.

The ANS controls the body’s stress response to various stressors professed by the brain, neutralizes the stressors’ effects, and restores homeostasis [52]. The HRV is a symbolic signal for the evaluation of autonomic functions of the

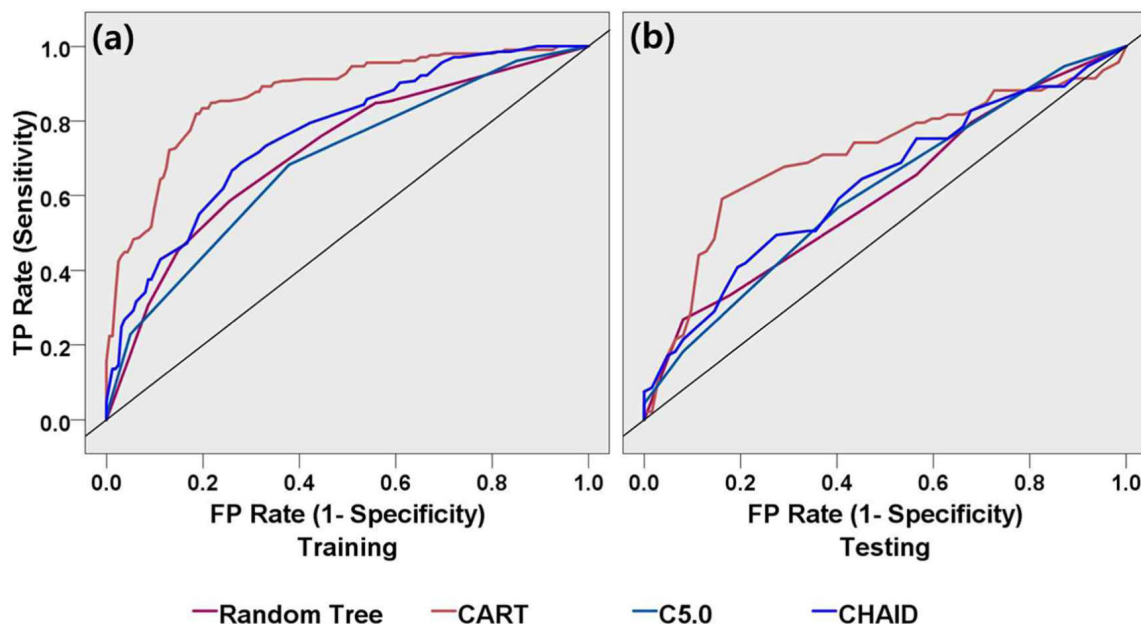


FIGURE 9. Receiver Operating Characteristic (ROC) curves for five different machine-learning models (Random Tree, CART, C5.0, CHAID). Area under ROC curve (AUC) is an indicator of prediction accuracy. (a) ROC curve of the training dataset. CART classified the training dataset with the highest AUC (87%) and highest accuracy (ACC: 82%); (b) ROC curve of the testing dataset. CART classified the testing dataset with the highest AUC (70%) and moderate accuracy (ACC: 69%). Diagonal black line is the reference line.

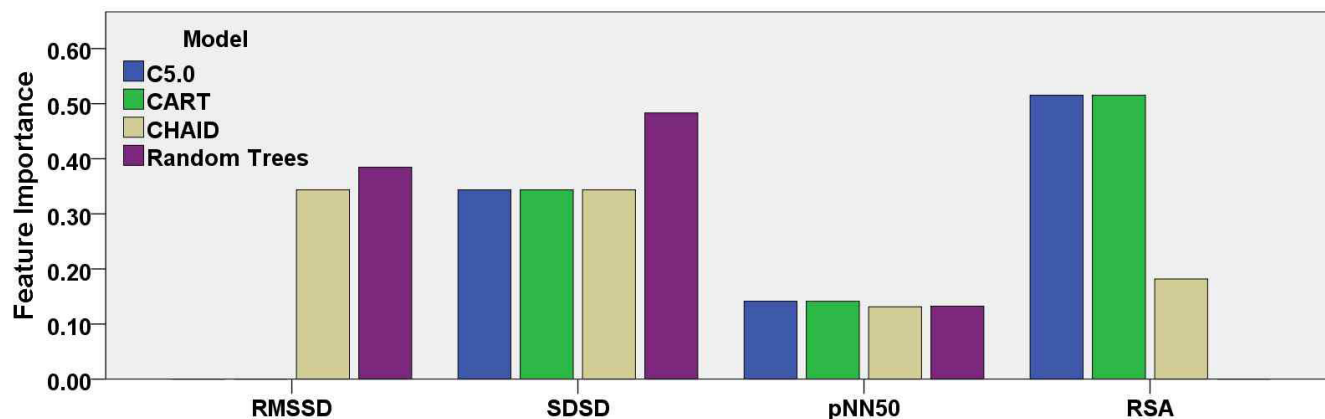


FIGURE 10. Feature importance of the HRV features in the feature ranking process of machine learning models to classify the stroke and control groups.

body. HRV features are clinically used as biomarkers for understanding the ANS changes after stroke [53]. As shown in Table 7, HRV characteristics, such as LF / HF, LF ratio, and HF ratio showed a strong association with the disturbed autonomic function derived from the stroke [45], [46], [54]. In this study, LF/HF, LF/(LF + HF), and HF/(LF + HF) have shown significant differences in the stroke group relative to control group. HRV time-domain features, such as RMSSD, RR-I, SDDSD have come out as the most predictive autonomic features for higher stroke risk [43]. This study revealed RMSSD, RSA, SDDSD, and pNN50 as the distinctive features during post-stroke treatment. Autonomic dysfunction is evident in the impaired physiological regulation of heart rate and increased cortisol secretion [3]. As cardiosympathetic centers are assumed in the anterior, medial, and

superior sections of the insula, stroke lesion in the inferior parietal and posterior insula may impair the parietal lobe’s link with autonomic centers causes an autonomic imbalance and increased risk of cardiac events [55]. HF power represents parasympathetic activity, and LF power correlates with vagal activity. The LF/HF, an indicator of sympathovagal balance, is significantly lower among the stroke patients than among the healthy control adults. Previous findings supported this study, revealing that higher HF power, lower LF power, and reduced LF/HF ratio predict post-stroke sub-acute infections [53], [56] and poor neurological outcomes [45]. Machine learning approaches enabled early stroke prognostics and most-stroke recovery using the cardiac activity profile. In our study, the decision tree-based Random Trees Model most accurately classified the stroke impaired cardiac

TABLE 7. Comparison of results of ECG-derived cardiac features of stroke population between proposed work and previous works.

Study	Study population	Purpose	Primary Results
Fyfe-Johnson et al.[43]	12,550 healthy adults, including 816 stroke patients	To evaluate the association between HRV and stroke	Individuals in the lowest quintiles of SDNN, mean NN, and RMSSD were at a higher risk of stroke.
Korpelainen et al. [44]	46 stroke patients and 30 healthy controls	To assess the effects of brain infarction on the HRV	SDNN, VLF, LF, and SD2 significantly reduced in stroke patients compared with the healthy controls
Graff et al.[45]	75 ischemic strokes	To access HRV as a prognostic tool of stroke.	Elevated HF, decreased LF, decreased LF/HF predictor of poor neurological outcome at 90 days
Wei et al. [46]	232 patients with acute ischemic stroke	To evaluate the association between autonomic function and stroke	SDNN, VLF, LF, and LF/HF derived from a 12-channel 24-hour ambulatory ECG differed significantly between the stroke patient groups (wand healthy control
This study	45 stroke patients and 40 healthy control volunteers	To access the feasibility of ECG as a prognostic tool for stroke prediction.	RR-I, QT, ST, QRS, SDSA, LF/HF, LF/(LF+HF), and HF/(LF+HF) are significantly different in stroke group relative to the control group.

profiles. We found ECG fiducial profiles as key predictive features for stroke prediction. According to statistical analysis, we found that most of the fiducial features are statistically significant. In contrast, only a few HRV features showed significant differences in discriminating the stroke group and the healthy group. These findings reveal that ECG fiducial features are more reliable to distinguish the stroke group and the healthy group. According to the machine-learning approach, we found that classification of ML model using fiducial features resulted in higher accuracy relative to HRV features. Therefore, our results demonstrated that fiducial features are more accurate predictors to classify the stroke group and the healthy group. In the future, we will try to explore ECG data recorded within a few weeks after the stroke onset to investigate whether the near stroke period HRV shows better classification performance than studied HRV data within three months. We have a plan to examine the combined performance of both features in the future. In Table 8 A, a comparison of classification performance of several machine-learning models were demonstrated for stroke prediction using ECG derived cardiac features.

To the best of our knowledge, the Big-ECG we developed was the first to propose an outpatient ECG-based

TABLE 8. Comparison of classification performance results of machine-learning models for stroke prediction in the proposed work and previous works.

Study	Study population	Purpose	Primary Results
Khosla et al.[57]	4,988 ECG recording with 299 occurrences of stroke from Cardiovascular Health Study (CHS) dataset	To evaluate the association between ECG and stroke	Margin-based Censored Regression (MCR) Model: AUC= 0.777 Concordance Index: 0.770
Attia et al. [58]	6,49,931 normal sinus rhythm ECGs from 1,80,922 patients	To detect ECG Atrial Fibrillation (AF) present during normal sinus rhythm as a biomarker for stroke.	Convolutional neural network (CNN) Model: AUC= 0.90 Sensitivity= 82.3%, ACC= 83.3%
Ledezma et al. [59]	Simulated ECG Modeling for ischemic stroke	To classify ischemic stroke using Artificial Neural Networks (ANN)	ANN Model: Sensitivity and positive predictive value above 95%
Shen et al. [33]	70 overnight single-lead ECG from apnea-ECG database	To classify obstructive sleep apnea (OSA) using deep-learning model.	Multiscale dilation attention CNN (MSDA-1DCNN) and weighted-loss time-dependent (WLTD) model:, Accuracy 89.4%, sensitivity 89.8%, and specificity: 89.1%
This study	45 stroke patients and 40 healthy control	To access the feasibility of ECG as a prognostic tool for stroke prediction.	Random Trees Model: (AUC: 98.9, overall accuracy: 95.6%).

cyber-physical system for managing stroke prognosis and post-stroke treatment. Several studies in the past have used standard 12-lead ECG with standalone devices. Real-time healthcare service in a non-clinical environment, such as rest, sleep, demands an ambulatory ECG along with instant data processing and health analytics. Thus, wearable ECG sensors, big-data-driven cloud analytics, and real-time service dashboards improve stroke prognosis and post-stroke rehabilitation management. Our system can be a useful measure to predict wake-up strokes in night sleep settings. This proposed ECG-based cyber-physical system can be a prospective HRV based sleep quality and sleep disorder monitoring system.

In this study, our focus belongs only to the single-channel ECG to understand the changes in the ECG for cardiac complications from ischemic stroke, not all standard ECG 12-leads. Lead V5 is identical to other lead positions, but there are still specific cardiac outcomes at each lead

position. For this reason, the model developed here is currently only generalized to Lead V5 ECG of stroke patients through current parameterization. We utilized 5 minutes ECG data recorded once, within 3 months after the stroke onset, long ECG changes were not studied in this study. As multiple features were extracted from each subject and leave-k-subject-out cross-validation was not performed, there is a possibility of potential subject-bias in cross-validation of this studies. Multimodal physiological data (EEG, PPG, EMG) may enhance the prediction accuracy of stroke-derived neural, vascular, and postural impairment in the cost of the computational power. Although the proposed cyber-physical system demonstrated ECG-based patient management, it is possible to integrate multiple physiological sensors, such as the EEG and PPG sensors, to monitor patients in various physiological domains. In the future, we will extend our system with a multimodal physiological sensing system for automated stroke prognosis and post-stroke rehabilitation studies. Moreover, leave-k-subject-out cross-validation will be performed to avoid subject-bias in future studies.

VII. CONCLUSION

Big-ECG, a cyber-physical cardiac monitoring system, was constructed for the stroke prognosis and post-stroke patient monitoring. We explained the sensor system, data analysis of the big data platform, and machine learning-based stroke prediction in detail. We successfully perform data acquisition, cloud-based data transformation, disease prediction, and visualization of 45 stroke patients and 40 healthy volunteers using this cyber-physical system. RR-I, QT, ST, QRS, SDSA, LF/HF, LF/(LF + HF), and HF/(LF + HF) were statistically significant cardiovascular biomarkers for identifying cardiac changes derived from an ischemic stroke during the post-stroke rehabilitation. The Big-ECG system is likely to be a prospective medical support system for the prognosis of ischemic stroke and post-stroke recovery.

REFERENCES

- [1] *World Health Statistics 2016: Monitoring Health for the SDGs Sustainable Development Goals*, World Health Org., Geneva, Switzerland, 2016.
- [2] 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.
- [3] P. Sörös and V. Hachinski, "Cardiovascular and neurological causes of sudden death after ischaemic stroke," *Lancet Neurol.*, vol. 11, no. 2, pp. 179–188, Feb. 2012.
- [4] I. Hussain and S.-J. Park, "Prediction of myoelectric biomarkers in post-stroke gait," *Sensors*, vol. 21, no. 16, p. 5334, Aug. 2021.
- [5] I. Hussain and S.-J. Park, "Quantitative evaluation of task-induced neurological outcome after stroke," *Brain Sci.*, vol. 11, no. 7, p. 900, Jul. 2021.
- [6] I. Hussain and S. J. Park, "HealthSOS: Real-time health monitoring system for stroke prognostics," *IEEE Access*, vol. 8, pp. 213574–213586, 2020.
- [7] S. Wu, D. Liang, Q. Yang, and G. Liu, "Regularity of heart rate fluctuations analysis in obstructive sleep apnea patients using information-based similarity," *Biomed. Signal Process. Control*, vol. 65, Mar. 2021, Art. no. 102370.
- [8] W. Pan, A. He, K. Feng, Y. Li, D. Wu, and G. Liu, "Multi-frequency components entropy as novel heart rate variability indices in congestive heart failure assessment," *IEEE Access*, vol. 7, pp. 37708–37717, 2019.
- [9] T. Lees, F. Shad-Kaneez, A. M. Simpson, N. T. Nassif, Y. Lin, and S. Lal, "Heart rate variability as a biomarker for predicting stroke, post-stroke complications and functionality," *Biomark Insights*, vol. 13, pp. 1–13, Jun. 2018.
- [10] K.-O. Lövsblad, S. Altrichter, V. M. Pereira, M. Vargas, A. M. Gonzalez, S. Haller, and R. Sztajzel, "Imaging of acute stroke: CT and/or MRI," *J. Neuroradiol.*, vol. 42, no. 1, pp. 55–64, Feb. 2015.
- [11] R. A. Macdonell, G. A. Donnan, P. F. Bladin, S. F. Berkovic, and C. Wriedit, "The electroencephalogram and acute ischemic stroke: Distinguishing cortical from lacunar infarction," *Arch. Neurol.*, vol. 45, no. 5, pp. 520–524, 1988.
- [12] S. J. Park, S. Hong, D. Kim, I. Hussain, Y. Seo, and M. K. Kim, "Physiological evaluation of a non-invasive wearable vagus nerve stimulation (VNS) device," in *Advances in Human Factors in Wearable Technologies and Game Design*. Cham, Switzerland: Springer, 2020, pp. 57–62.
- [13] R. Baheti and H. Gill, "Cyber-physical systems," *Impact Control Technol.*, vol. 12, no. 1, pp. 161–166, 2011.
- [14] S. J. Park, S. Hong, D. Kim, I. Hussain, and Y. Seo, "Intelligent in-car health monitoring system for elderly drivers in connected car," in *Proc. 20th Congr. Int. Ergonom. Assoc. (IEA)*. Cham, Switzerland: Springer, 2018, pp. 40–44.
- [15] G. Yang, Z. Pang, M. Jamal Deen, M. Dong, Y.-T. Zhang, N. Lovell, and A. M. Rahmani, "Homecare robotic systems for healthcare 4.0: Visions and enabling technologies," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 9, pp. 2535–2549, Sep. 2020.
- [16] D. Chen, Y. Chen, B. N. Brownlow, P. P. Kanjamala, C. A. G. Arredondo, B. L. Radspinner, and M. A. Raveling, "Real-time or near real-time persisting daily healthcare data into HDFS and elasticsearch index inside a big data platform," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 595–606, Apr. 2017.
- [17] A. Abouzeid, K. Bajda-Pawlikowski, D. Abadi, A. Silberschatz, and A. Rasin, "HadoopDB: An architectural hybrid of MapReduce and DBMS technologies for analytical workloads," *Proc. VLDB Endowment*, vol. 2, no. 1, pp. 922–933, Aug. 2009.
- [18] S. Hong, D. Kim, H. Park, Y. Seo, H. Iqram, and S. Park, "Gait feature vectors for post-stroke prediction using wearable sensor," *Korean Soc. Emotion Sensibility*, vol. 22, no. 3, pp. 55–64, Sep. 2019.
- [19] S. J. Park, S. Hong, D. Kim, Y. Seo, and I. Hussain, "Knowledge based health monitoring during driving," in *HCI International 2018—Posters' Extended Abstracts*. Cham, Switzerland: Springer, 2018, pp. 387–392.
- [20] S. J. Park, I. Hussain, S. Hong, D. Kim, H. Park, and H. C. M. Benjamin, "Real-time gait monitoring system for consumer stroke prediction service," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2020, pp. 1–4.
- [21] S. J. Park, S. Hong, D. Kim, Y. Seo, I. Hussain, J. H. Hur, and W. Jin, "Development of a real-time stroke detection system for elderly drivers using quad-chamber air cushion and IoT devices," *SAE Tech. Paper 0148-7191*, 2018.
- [22] S. J. Park, S. Hong, D. Kim, J. H. Yu, I. Hussain, H. G. Park, and H. C. M. Benjamin, "Development of intelligent stroke monitoring system for the elderly during sleeping," *Sleep Med.*, vol. 64, p. S294, Dec. 2019.
- [23] I. Hussain, S. Young, C. H. Kim, H. C. M. Benjamin, and S. J. Park, "Quantifying physiological biomarkers of a microwave brain stimulation device," *Sensors*, vol. 21, no. 5, p. 1896, Mar. 2021.
- [24] B. Snyder, D. Bosnanac, and R. Davies, *ActiveMQ in Action*. Greenwich, CT, USA: Manning Greenwich Connecticut, 2011.
- [25] C. Gormley and Z. Tong, *Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine*. Sebastopol, CA, USA: O'Reilly Media, 2015.
- [26] R. Studer, V. R. Benjamins, and D. Fensel, "Knowledge engineering: Principles and methods," *Data Knowl. Eng.*, vol. 25, nos. 1–2, pp. 161–197, 1998.
- [27] M. A. Musen and T. Protégé, "The Protégé project: A look back and a look forward," *AI Matters*, vol. 1, no. 4, pp. 4–12, 2015.
- [28] G. Liu, Q. Wang, S. Chen, G. Zhou, W. Chen, and Y. Wu, "Robustness evaluation of heart rate variability measures for age gender related autonomic changes in healthy volunteers," *Australas. Phys. Eng. Sci. Med.*, vol. 37, no. 3, pp. 567–574, Sep. 2014.
- [29] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.
- [30] D. Kim, S. Hong, I. Hussain, Y. Seo, and S. J. Park, "Analysis of bi-signal data of stroke patients and normal elderly people for real-time monitoring," in *Proc. 20th Congr. Int. Ergonom. Assoc. (IEA)*. Cham, Switzerland: Springer, 2019, pp. 208–213.
- [31] H. Park, S. Hong, I. Hussain, D. Kim, Y. Seo, and S. J. Park, "Gait monitoring system for stroke prediction of aging adults," in *Advances in Human Factors in Wearable Technologies and Game Design*. Cham, Switzerland: Springer, 2020, pp. 93–97.

- [32] S. Chen, K. Xu, X. Yao, S. Zhu, B. Zhang, H. Zhou, X. Guo, and B. Zhao, "Psychophysiological data-driven multi-feature information fusion and recognition of miner fatigue in high-altitude and cold areas," *Comput. Biol. Med.*, vol. 133, Jun. 2021, Art. no. 104413.
- [33] Q. Shen, H. Qin, K. Wei, and G. Liu, "Multiscale deep neural network for obstructive sleep apnea detection using RR interval from single-lead ECG signal," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.
- [34] M. Stone, "Cross-validators choice and assessment of statistical predictions," *J. Roy. Stat. Soc., B, Methodol.*, vol. 36, no. 2, pp. 111–133, Jan. 1974.
- [35] G. V. Kass, "An exploratory technique for investigating large quantities of categorical data," *J. Roy. Stat. Soc., C, Appl. Statist.*, vol. 29, no. 2, pp. 119–127, 1980.
- [36] J. R. Quinlan. (2004). *Data Mining Tools See5 and C5.0*. [Online]. Available: <http://www.rulequest.com/see5-info.html>
- [37] W.-Y. Loh and Y.-S. Shih, "Split selection methods for classification trees," *Statist. Sinica*, vol. 7, no. 4, pp. 815–840, 1997.
- [38] C. M. Bishop, *Neural Networks for Pattern Recognition*. London, U.K.: Oxford Univ. Press, 1995.
- [39] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*. Boca Raton, FL, USA: CRC Press, 1984.
- [40] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 3, pp. 299–310, Mar. 2005.
- [41] S. Kumar, M. H. Selim, and L. R. Caplan, "Medical complications after stroke," *Lancet Neurol.*, vol. 9, no. 1, pp. 105–118, Jan. 2010.
- [42] J. Prosser, L. MacGregor, K. R. Lees, H.-C. Diener, W. Hacke, and S. Davis, "Predictors of early cardiac morbidity and mortality after ischemic stroke," *Stroke*, vol. 38, no. 8, pp. 2295–2302, Aug. 2007.
- [43] A. L. Fyfe-Johnson, C. J. Muller, A. Alonso, A. R. Folsom, R. F. Gottesman, W. D. Rosamond, E. A. Whitsel, S. K. Agarwal, and R. F. MacLehose, "Heart rate variability and incident stroke: The atherosclerosis risk in communities study," *Stroke*, vol. 47, no. 6, pp. 1452–1458, 2016.
- [44] J. T. Korpelainen, K. A. Sotaniemi, A. Mäkikallio, H. V. Huikuri, and V. V. Myllylä, "Dynamic behavior of heart rate in ischemic stroke," *Stroke*, vol. 30, no. 5, pp. 1008–1013, May 1999.
- [45] B. Graff, D. Gąsecki, A. Rojek, P. Boutouyrie, W. Nyka, S. Laurent, and K. Narkiewicz, "Heart rate variability and functional outcome in ischemic stroke: A multiparameter approach," *J. Hypertension*, vol. 31, no. 8, pp. 1629–1636, 2013.
- [46] L. Wei, W.-B. Zhao, H.-W. Ye, Y.-H. Chen, X.-P. Zhang, Y. Huang, Y.-F. Cai, Q.-F. Chen, and S.-Y. Pan, "Heart rate variability in patients with acute ischemic stroke at different stages of renal dysfunction: A cross-sectional observational study," *Chin. Med. J.*, vol. 130, no. 6, p. 652, 2017.
- [47] M. A. Samuels, "The brain-heart connection," *Circulation*, vol. 116, no. 1, pp. 77–84, 2007.
- [48] 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.
- [49] M. M. McDermott, F. Lefevre, M. Arron, G. J. Martin, and J. Biller, "ST segment depression detected by continuous electrocardiography in patients with acute ischemic stroke or transient ischemic attack," *Stroke*, vol. 25, no. 9, pp. 1820–1824, Sep. 1994.
- [50] K. Y. K. Wong, R. S. M. Walter, D. Douglas, H. W. Fraser, S. A. Ogston, and A. D. Struthers, "Long QTc predicts future cardiac death in stroke survivors," *Heart*, vol. 89, no. 4, p. 377, 2003.
- [51] L. G. Stead, R. M. Gilmore, M. F. Bellolio, L. Vaidyanathan, A. L. Weaver, W. W. Decker, and R. D. Brown, "Prolonged QTc as a predictor of mortality in acute ischemic stroke," *J. Stroke Cerebrovascular Diseases*, vol. 18, no. 6, pp. 469–474, Nov. 2009.
- [52] J. M. Koolhaas, A. Bartolomucci, B. Buwalda, S. F. de Boer, G. Flügge, S. M. Korte, P. Meerlo, R. Murison, B. Olivier, P. Palanza, G. Richter-Levin, A. Sgoifo, T. Steimer, O. Stiedl, G. van Dijk, M. Wöhr, and E. Fuchs, "Stress revisited: A critical evaluation of the stress concept," *Neurosci. Biobehav. Rev.*, vol. 35, no. 5, pp. 1291–1301, Apr. 2011.
- [53] A. Günther, I. Salzmann, S. Nowack, M. Schwab, R. Surber, H. Hoyer, O. W. Witte, and D. Hoyer, "Heart rate variability—A potential early marker of sub-acute post-stroke infections," *Acta Neurol. Scandinavica*, vol. 126, no. 3, pp. 189–196, Sep. 2012.
- [54] J. T. Korpelainen, K. A. Sotaniemi, H. V. Huikuri, and V. V. Myllylä, "Circadian rhythm of heart rate variability is reversibly abolished in ischemic stroke," *Stroke*, vol. 28, no. 11, pp. 2150–2154, Nov. 1997.
- [55] F. Rincon, M. Dhamoon, Y. Moon, M. C. Paik, B. Boden-Albala, S. Homma, M. R. Di Tullio, R. L. Sacco, and M. S. V. Elkind, "Stroke location and association with fatal cardiac outcomes," *Stroke*, vol. 39, no. 9, pp. 2425–2431, Sep. 2008.
- [56] S. M. Oppenheimer, G. Kedem, and W. M. Martin, "Left-insular cortex lesions perturb cardiac autonomic tone in humans," *Clin. Autonomic Res.*, vol. 6, no. 3, pp. 131–140, Jun. 1996.
- [57] A. Khosla, Y. Cao, C. C.-Y. Lin, H.-K. Chiu, J. Hu, and H. Lee, "An integrated machine learning approach to stroke prediction," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2010, pp. 183–192.
- [58] Z. I. Attia, P. A. Noseworthy, F. Lopez-Jimenez, S. J. Asirvatham, A. J. Deshmukh, B. J. Gersh, R. E. Carter, X. Yao, A. A. Rabinstein, B. J. Erickson, S. Kapa, and P. A. Friedman, "An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction," *Lancet*, vol. 394, no. 10201, pp. 861–867, Sep. 2019.
- [59] C. A. Ledezma, X. Zhou, B. Rodriguez, P. J. Tan, and V. Dfaz-Zuccarini, "A modeling and machine learning approach to ECG feature engineering for the detection of ischemia using pseudo-ECG," *PLoS ONE*, vol. 14, no. 8, Aug. 2019, Art. no. e0220294.



IQRAM HUSSAIN (Graduate Student Member, IEEE) received the B.Sc. degree in mechanical engineering from Khulna University of Engineering & Technology, Bangladesh. He is currently pursuing the Ph.D. degree with the Department of Medical Physics, University of Science and Technology (UST), South Korea. He is also a Research Associate with the Center for Medical Convergence Metrology, Korea Research Institute of Standards and Science (KRISS), Daejeon, South Korea. Besides, he works in Knowledgebase Super Brain (KSB) Project with the Electronics and Telecommunication Research Institute (ETRI), Daejeon. At the beginning of his professional career, he worked with Summit Power International Ltd., and Doreen Power Generations and Systems Ltd., in the electricity generation plants, Dhaka, Bangladesh. He has ten years of work experience in power plant operation and maintenance, and power plant project management. His current research interests include wearable sleep monitoring, big-data analytics, neuroscience, medical physics, brain-computer interaction, and ergonomics.



SE JIN PARK received the Ph.D. degree in industrial engineering from Korea University, in 1994. He is currently the Director of the Data Center for Korean Body Measurement, Korea Research Institute of Standards and Science (KRISS), supported by the Ministry of Trade, Industry and Energy, South Korea. Since joining KRISS, in 1988, he has served in various positions, including the Director of convergence technology, the Head of ergonomics research, and the Head of medical metrology. He is also the CEO of Sewon Intelligence, an AI-based healthcare company, Seoul, South Korea. He is also working as a Researcher with the Electronics and Telecommunications Research Institute (ETRI), South Korea. Besides, he is also a Professor with the Department of Medical Physics, University of Science and Technology (UST), South Korea. He is also a Public Speaker and provides speeches in electronic media on a wide range of issues. His research interests include human factors and ergonomics, biomechanics, emotion and sensibility, human vibration and seating comfort, human-computer interaction (HCI), Internet of Everything (IoE), and gerontechnology. He served as the President of Korean Society of Emotion and Sensibility (KOSES) and the Ergonomics Society of Korea.