**AI-Based Diabetes based Prediction System**

**1. Introduction**

**The Fourth Industrial Revolution has arrived, bringing with it immense benefits as well as many challenges for various industries and research fields. In the healthcare field, the focus has been on information and communication technologies (ICT) such as artificial intelligence, big data, the Internet of Things (IoT), and cloud computing, which are also the core technologies for this Fourth Industrial Revolution [1–4]. For example, IoT enables the exchange of different types of health and medical information (bio-signals, past medical history, and genetic data) between various medical devices and medical institution systems. According to the WHO, the world’s population is rapidly aging [5]; this will lead to increases in chronic diseases as well as healthcare costs. In anticipation of this, countries are shifting the focuses of their healthcare systems from sickness and disease to prevention and wellness. In addition, health information such as personal health records (PHR), electronic medical records (EMR), and genomic information is continually being generated, collected, and stored, and it can be easily used for analysis by making it big data. Over the years, a vast amount of medical data has been generated, collected, and stored, but it has yet to be properly used. An in-depth analysis of these data that combines big data and artificial intelligence (AI) technologies can help develop new** intelligent medical solutions, such as precision healthcare and predictive healthcare services, that can be used to prevent diseases. However, it is still difficult to derive meaning information between various types of healthcare big data. The recent improvements in computing infrastructure and the emergence of various AI frameworks in ICT have made AI-based digital healthcare analysis more intelligent and more feasible for this smart healthcare era. The smart healthcare system is evolving into a service that combines medical big data with ICT to help individuals manage their health remotely.

**2. Related Works:**

**Concept of Stroke** :

**Stroke is one of the major diseases associated with death worldwide, and it causes cognitive and functional disorders [5,19,20]. A stroke happens when the blood vessels in the brain either become clogged or burst, thus reducing oxygen supply to the brain cells, resulting in necrosis of brain tissue [6,19]. As cells in the brain die, certain parts of the body lose functionality.**

**This leads to various symptoms such as loss of body coordination, and speech and sensory impairment. Stroke can be categorized into ischemic stroke (cerebral infarction), which is caused by blockages in the blood vessels, and cerebral hemorrhagic stroke (cerebral hemorrhage), which is caused by the rupturing of blood vessels in the brain [20]. Cerebral infarction diseases can be divided into cerebrovascular thrombosis and cerebral embolism.**

**Cerebral thrombosis is a symptom that is caused by blocking blood clots in the brain due to arteriosclerosis or having problems with the inner wall of the blood vessel. While cerebral embolism is a condition caused by blood clots from the heart which blocks the blood vessels bring oxygen and blood to the brain. Next, there are two types of hemorrhagic stroke: intracerebral hemorrhage and subarachnoid hemorrhage. An intracerebral hemorrhage causes weak blood vessels to burst if there is a sudden rise in blood pressure. Brain cells that are supplied oxygen and nutrients by the arteries that have burst in turn become damaged, and the surrounding cells are crushed by the burst blood.**

**Hypertension has been reported to be the main cause of most of these symptoms of intracerebral hemorrhage. Meanwhile, subarachnoid hemorrhage is caused by ruptured intracranial aneurysm irritating the lining of the brain. Subarachnoidal hemorrhage can be divided into spontaneous hemorrhage and traumatic hemorrhage. 80% of subarachnoid hemorrhage is caused by the ruptured cerebral aneurysm, and it is usually suspected when there is subarachnoid hemorrhage. Symptoms of subarachnoid hemorrhage vary from sudden severe headache, severe nausea, vomiting to loss of consciousness. However, the most characteristic symptom is a sudden severe headache unlike anything the person has experienced before.**

**This subarachnoidal hemorrhage injury is reported to be fatal injury that cause death in one-third of patients before they can arrive at a hospital, and only the remaining patients are known to receive treatment [19,20]. According to Statistics Korea, the total number of deaths in Korea in 2018 was 298,820, with 161,187 for men and 137,633 for women [21].**

**The causes of death were reported to be 79,153 cases of malignant neoplasm (cancer), 32,004 cases of heart disease, 23,280 cases of pneumonia, and 22,940 cases of cerebrovascular disease. Cerebrovascular disease is the third-highest single disease cause of death, with high mortality rates of 42.7 for men and 46.7 for women per 100,000 [21].**

**Although the death toll from cerebrovascular disease has been declining since 2005, it is still a high-risk disease that ranks third among all single disease causes of death. In particular, for those aged 60 or older, the mortality rate from heart disease and cerebrovascular disease is gradually increasing.**

**Fast detection and treatment are paramount in the early onset of a stroke, as untreated stroke can leave severe aftereffects, such as hemiplegia or even death.**

**3. Artificial Intelligence-Based Stroke Disease Prediction System Using EMG**

**In this paper, we propose a new AI-based stroke disease system using EMG bio-signals from everyday life as shown in Figure 1.**

**The first module executes offline processing and includes the functions of machine learning and the deep learning-based learning model generation and management of EMG data.**

**The second module performs an online processing function as well as early detection and prediction of stroke based on EMG bio-signals collected in real-time from everyday life.**

**In the offline module, the EMG bio-signals data measured in real-time during daily walking is updated in the repository according to the cycle set by the system.**

**Preprocessing of the collected EMG bio-signals is performed, and a learning model is generated using machine learning and deep learning algorithms.**

**Attribute subset selection from the preprocessed EMG is used to generate learning models with machine learning algorithms for early detection and prediction of stroke diseases.**

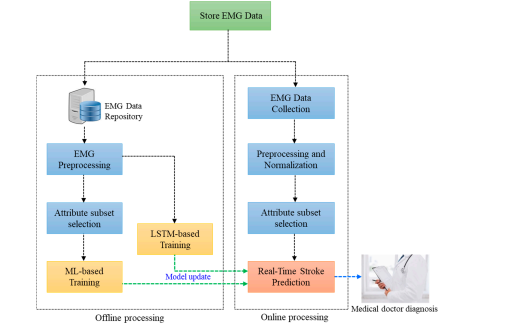
**Machine learning LSTM-based learning models are developed and sent for online processing to the second module before being used in real-time stroke predictions. The online module measures and collects real-time EMG bio-signals in daily life at the request of the system or user.**

**The “preprocessing and normalization” block removes missing or incomplete data from the collected EMG data. As the minimum and the maximum values are different for each attribute, a normalization process is performed depending on the measurement unit. In the attribute subset selection block, the prediction accuracy and the analysis speed are improved by selecting the EMG attribute subset defined in this paper.**

**In addition, it is possible to guarantee optimal performance of the predicted model learned and provide analytical information. The ‘Real-Time Stroke Prediction’ block is mounted with pre-trained machine learning and LSTM-based prediction models. Real-time prediction and semantic analysis based on machine learning are performed using the selected EMG optimal attributes subset.**

**At this time, the deep-learning LSTM model implements early detection and prediction of stroke in real-time based on EMG bio-signals that went through preprocessing blocks. These predictions and semantic analysis information are then transmitted to the medical staff or hospital.**

**Finally, based on the medical doctor’s diagnosis for the risk of stroke, the patients are provided assistance with receiving medical examinations and treatment services with emergency alarms and quick hospital visits as appropriate. Appl. Sci. 2020, 10, x FOR PEER REVIEW.**

****

**4. Experiments and Analysis:**

**Dataset and Experimental Analysis:**

**This section describes the process of measuring and collecting bio-signal data for verification of the proposed AI-based stroke disease prediction system.**

**The main bio-signal used will be real-time electromyography (EMG) data. EMG is a diagnostic tool that records electrical activity inside a specific muscle or measures nerve conduction velocity through electrical stimulation using electrodes. Comprehensive studies of strokes using EMG have shown that there is a slight imbalance in the body both before and after a stroke, along with imbalances in gait and locomotion [48,49].**

**In this paper, we analyze these imbalances and gait disorders in terms of EMG bio-signal data, which has the potential to be a stroke risk factor. The measurement and collection of EMG biometric data were conducted from 2015 to 2017 at the emergency medical center and the department of rehabilitation medicine at Chungnam National University Hospital.**

**Our stroke group consisted of patients aged 65 or older who were undergoing rehabilitation treatment for stroke within one month of stroke confirmation. In total, 287 patients from the rehabilitation department fit our criteria. Various biological signals were collected, such as EMG, ECG (electrocardiogram), EEG (electroencephalogram), foot pressure, and voice recordings. Figure 2 shows the location of each sensor on the patient. Before collecting bio-signal data through experimentation, the sensors were checked to confirm they were operating normally.**

**Patients who were undergoing rehabilitation other than a stroke were considered as the normal group, and data were collected from 271 individuals. Each patient was put through various scenarios such as standing, walking, sitting, raising arms, and sleeping to simulate everyday activities.**

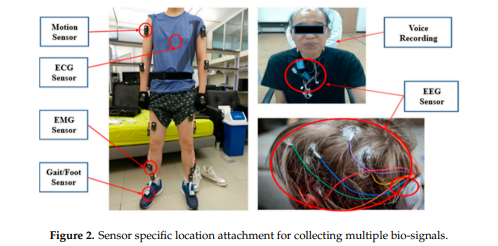
**For each scenario, all subjects had one practice before executing the measurement protocol. Despite this prior practice, the first measured and collected values were not used as experimental data, because human noise due to the subject’s tension and discomfort could be reflected.**

**At the same time, the last measurement protocol was also not reflected because repetitive experiments cause fatigue to elderly subjects. Each bio-signals data collection information was delivered to the relay in real time and the BLE (bluetooth low energy) protocol was used for communication. In addition, data was forwarded to the Wi-Fi communication protocol to the server that collects and predicts bio-signals from the gateways. At this time, the entire process of the measurement experiment was monitored by the medical doctor and re-measurement was performed if there was a loss of bio-signal data transmitted.**

**The raw data transmitted from the four locations of the EMG was set in the system to have a false value of 4 bytes per sampling rate in the form of a voltage. Appl. Sci. 2020, 10, x FOR PEER REVIEW 8 of 19 measurement protocol was also not reflected because repetitive experiments cause fatigue to elderly subjects.**

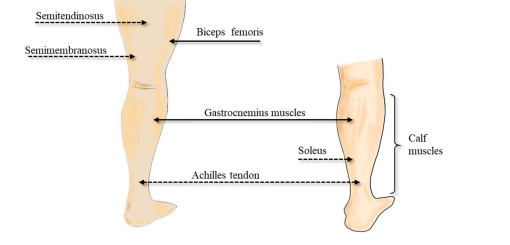
**Each bio-signals data collection information was delivered to the relay in real time and the BLE (bluetooth low energy) protocol was used for communication.**

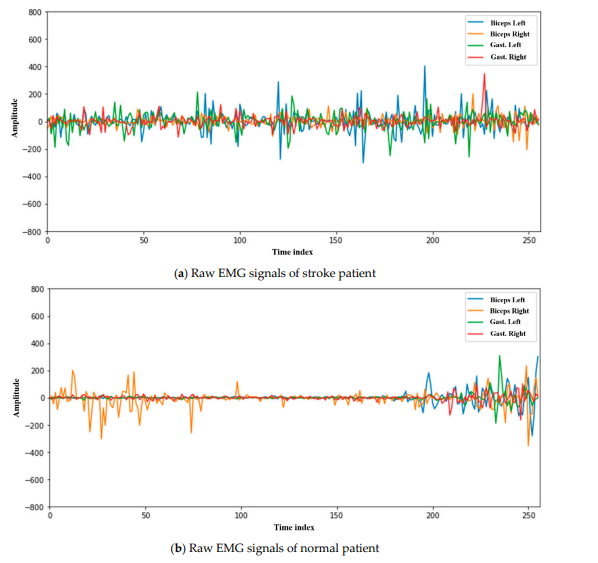
**In addition, data was forwarded to the Wi-Fi communication protocol to the server that collects and predicts bio-signals from the gateways. At this time, the entire process of the measurement experiment was monitored by the medical doctor and re-measurement was performed if there was a loss of bio-signal data transmitted. The raw data transmitted from the four locations of the EMG was set in the system to have a false value of 4 bytes.**

****

**Figure 3 shows the measurement locations used to collect the EMG bio-signals. EMG signals were collected with a sampling rate of 1500 Hz per second from a total of four locations: the left and right legs, biceps femoris, and gastrocnemius muscle.**

**As a stroke patient and a normal person should be classified based on LSTM of machine learning and a deep learning model, it is necessary to ensure an equal amount of experimental data for the two classes for predicting a stroke. Therefore, 271 out of 287 stroke patients were randomly extracted, and 271 normal subjects’ data were used. Figure 4 shows raw EMG signals for a random stroke patient and a normal patient collected at the four muscle locations while walking, which will be used in the model training and the testing of LSTM.**

****

****

**5. Conclusions :**

**We presented a system using the Random Forest algorithm of machine learning and LSTM of deep learning that can detect and predict stroke based on real-time EMG bio-signals. The proposed system can significantly minimize the social and economic losses from stroke by predicting stroke in real-time. In this paper, four points of EMG raw data were measured and collected in real-time from Appl. Sci. 2020, 10, 6791 16 of 19 healthcare devices.**

**From this raw data, attributes were extracted and used with ML/DL models for stroke prediction accuracy and verification of this system. In addition, our experiment has shown that it is possible to detect and predict stroke symptoms with only bio-signals generated in everyday activities.**

**This system overcomes the limitations of other systems by providing the probability of the occurrence of a disease in the next 10 years, or the degree of severity after the outbreak. This prediction model is significant as it can reduce misdiagnosis levels and be used to warn medical staff or hospitals to preemptively respond to the health needs of older people through early detection and prediction of stroke diseases.**

**Additionally, the proposed Random Forest algorithm and LSTM-based stroke disease prediction model have the advantage of being able to be extended and applied as an early prediction model for diseases such as heart disease. In addition, we will develop a system that can analyze various type of data (diagnostic experience, medical knowledge, biometric signal analysis information, and EMR, etc.) in a scientific and comprehensive way so that it can directly help the diagnosis and decision of the medical doctors. Future research tasks include measuring and collecting real-time bio-signals in driving or sleeping services as well as walking during everyday life, along with conducting research and development on the provision of more comprehensive forecasting services for stroke diseases.**

**Author Contributions: Conceptualization, J.Y. and C.-S.P.; methodology, J.Y., H.L., C.-S.P., S.P., S.-H.K. and C.M.B.H.; software, J.Y., C.M.B.H.; validation, J.Y., H.L., C.-S.P., S.P., S.-H.K. and C.M.B.H.; formal analysis, J.Y., H.L., C.-S.P., S.P., S.-H.K. and C.M.B.H.; investigation, J.Y. and C.-S.P.; resources, J.Y. and C.-S.P.; data curation, J.Y. and H.L.; writing—original draft preparation, J.Y., H.L., C.-S.P., S.P., S.-H.K. and C.M.B.H.; writing—review and editing, J.Y. and H.L.; visualization, J.Y.; supervision, J.Y., H.L. and C.-S.P.; project administration, C.-S.P.; funding acquisition, C.-S.P. All authors have read and agreed to the published version of the manuscript. Funding: This work was supported by the National Research Council of Science & Technology (NST) grant by the Korea government (MSIP) (No. CRC-15-05-ETRI). Conflicts of Interest: The authors declare no conflict of interest.**

**Report:**

**All the above instructions are installed and executed successfully.**

**Project bt :**

**Name: S.VELAN**

**Dept/year: CSE/ III - YEAR**

**Reg No: 621421104056**

**College code:6214**

**Group: IBM-GROUP 5**