Artificial Intelligence-Based Speech Analysis System for Medical Support

Eui-Sun Kim[1], Dong Jin Shin[2], Sung Tae Cho[3], Kyung Jin Chung[4]

[1]Department of Media, Soongsil University, Seoul, Korea
[2]Department of Neurology, Gachon University Gil Medical Center, Incheon, Korea
[3]Department of Urology, Hallym University Kangnam Sacred Heart Hospital, Hallym University College of Medicine, Seoul, Korea
[4]Department of Urology, Gachon University Gil Medical Center, Incheon, Korea

Purpose: Prior research has indicated that stroke can influence the symptoms and presentation of neurogenic bladder, with various patterns emerging, including abnormal facial and linguistic characteristics. Language patterns, in particular, can be easily recognized. In this paper, we propose a platform that accurately analyzes the voices of stroke patients with neurogenic bladder, enabling early detection and prevention of the condition.

Methods: In this study, we developed an artificial intelligence-based speech analysis diagnostic system to assess the risk of stroke associated with neurogenic bladder disease in elderly individuals. The proposed method involves recording the voice of a stroke patient while they speak a specific sentence, analyzing it to extract unique feature data, and then offering a voice alarm service through a mobile application. The system processes and classifies abnormalities, and issues alarm events based on analyzed voice data.

Results: In order to assess the performance of the software, we first obtained the validation accuracy and training accuracy from the training data. Subsequently, we applied the analysis model by inputting both abnormal and normal data and tested the outcomes. The analysis model was evaluated by processing 30 abnormal data points and 30 normal data points in real time. The results demonstrated a high test accuracy of 98.7% for normal data and 99.6% for abnormal data.

Conclusions: Patients with neurogenic bladder due to stroke experience long-term consequences, such as physical and cognitive impairments, even when they receive prompt medical attention and treatment. As chronic diseases become increasingly prevalent in our aging society, it is essential to investigate digital treatments for conditions like stroke that lead to significant sequelae. This artificial intelligence-based healthcare convergence medical device aims to provide patients with timely and safe medical care through mobile services, ultimately reducing national social costs.

Keywords: Stroke; Speech recognition; Deep learning; Diagnosis support system; Neurogenic bladder

INTRODUCTION

Neurogenic bladder refers to bladder overactivity caused by nerve damage in patients who have experienced a stroke and have been diagnosed with overactive bladder, as well as abnormalities in bladder storage function identified through dynamic

Corresponding author: Kyung Jin Chung (ORCID) https://orcid.org/0000-0003-0153-7814
Department of Urology, Gachon University, Gil Medical Center, Namdong-daero 774 beon-gil, Namdong-gu, Incheon, 21565 Korea
Email: kjchung@gilhospital.com
Submitted: June 14, 2023 / **Accepted after revision:** June 26, 2023

testing. Previous studies [1,2] have reported that stroke can influence the symptoms or disease presentation of neurogenic bladder. There are several patterns associated with this condition, including distinct facial and speech features that deviate from the norm, with speech patterns being relatively easy to identify. These speech patterns may involve incorrect pronunciation or unnatural intonation, and accurately detecting these patterns is essential for stroke prevention. In this paper, we propose a platform that can precisely analyze the speech of stroke patients with neurogenic bladder, enabling early detection and prevention of the disease.

Cardiovascular disease in the elderly has become a significant health concern, ranking second among all causes of mortality in Korea, a country with an aging population, and the associated social costs are increasing [3-11]. Furthermore, stroke sequelae can manifest in various stages, ranging from mild to severe, including motor dysfunction, cognitive dysfunction, speech dysfunction, and depressive disorders. To prevent these diverse stroke sequelae, it is essential to develop a deep learning-based stroke alert software that can reduce the number of patients with chronic diseases.

Digital therapeutics, a subset of digital healthcare technologies, are poised to play a crucial role in the future of healthcare services by preventing, managing, and treating diseases rather than merely providing healthcare. Due to their software-based nature, digital therapeutics exhibit lower toxicity and fewer side effects compared to conventional treatments. Additionally, they do not require manufacturing, transportation, or storage like traditional medicines, allowing for easy large-scale distribution at a reduced cost, ultimately lowering medical expenses. This technology enables a small number of doctors to manage a large patient population, overcoming physical and temporal constraints, and potentially addressing issues such as health insurance financing, medical supply shortages, and regional disparities. In light of the coronavirus disease 2019 pandemic, digital therapeutics have garnered increased attention as they can partially replace in-person medical treatments for mental illnesses and chronic diseases, thereby reducing the risk of infection.

In the realm of stroke management, digital therapies — specifically for neurogenic bladder — are a form of software-based treatment that aims to modify patient behavior or lifestyle through various digital stimuli, such as images and sounds, while collecting and analyzing the resulting data. Presently, the application of digital therapeutics is progressively expanding into fields related to behavioral and habit changes, including mental illness and

chronic diseases. For instance, Akili Interactive Lab's Endeavor-Rx (AKL-T01, Akili Inc., Boston, MA, USA) is a video game designed to treat pediatric attention deficit hyperactivity disorder by selectively stimulating specific neural circuits. Sleepio, developed by Big Health (San Francisco, CA, USA), is an application that manages factors influencing sleep, such as negative thoughts, bedroom conditions, and lifestyle, in addition to sleep schedules, offering a 6-step treatment service. Moreover, digital therapy programs are being developed by companies like Voluntis and LifeSemantics for cancer treatment, Cognoa for pediatric behavior management, Omada Health, Noom, and Well-Doc for the prevention and management of diabetes and other chronic diseases, and MedRhythms for music therapy targeting motor, speech, and cognitive dysfunction.

Digital therapies can serve as a valuable complement to traditional medications in conditions that necessitate ongoing management. The software proposed in this paper has the potential to offer scientific monitoring feedback to users or nearby nursing staff by continuously learning from and analyzing the speech data of stroke patients, while also providing ongoing feedback and attention. Furthermore, it delivers therapeutic benefits for personalized stroke-related dysphonia associated with neurogenic bladder dysfunction. As such, this software is anticipated to function as an artificial intelligence (AI)-driven digital healthcare medical device.

MATERIALS AND METHODS

In this paper, we present the development of an AI-based speech analysis diagnostic system designed to identify the risk of stroke associated with neurogenic bladder disease in elderly individuals. Our proposed method involves recording the voice of a stroke patient as they speak a specific sentence, analyzing the recording to extract unique feature data, and then offering a voice alarm service through a mobile app. Fig. 1 below illustrates the overall system concept, which integrates stroke speech analysis software and clinical trial work.

Building Training Data
First, the data set is organized into a patient group and a control group. The patient group consisted of subjects who visited the hospital and were diagnosed with stroke, and the control group consisted of subjects who visited the hospital and were diagnosed as normal. The stroke diagnosis of the patient group is mild, characterized by initial facial paralysis and difficulty in
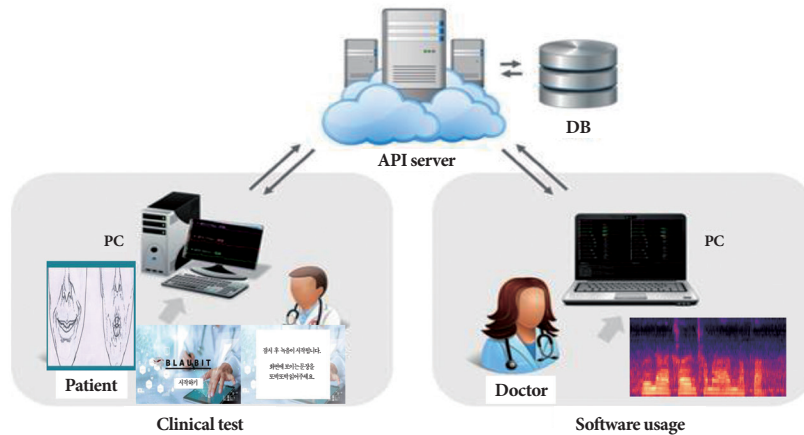
Fig. 1. Stroke speech analysis system overview. API, application programming interface; DB, database.
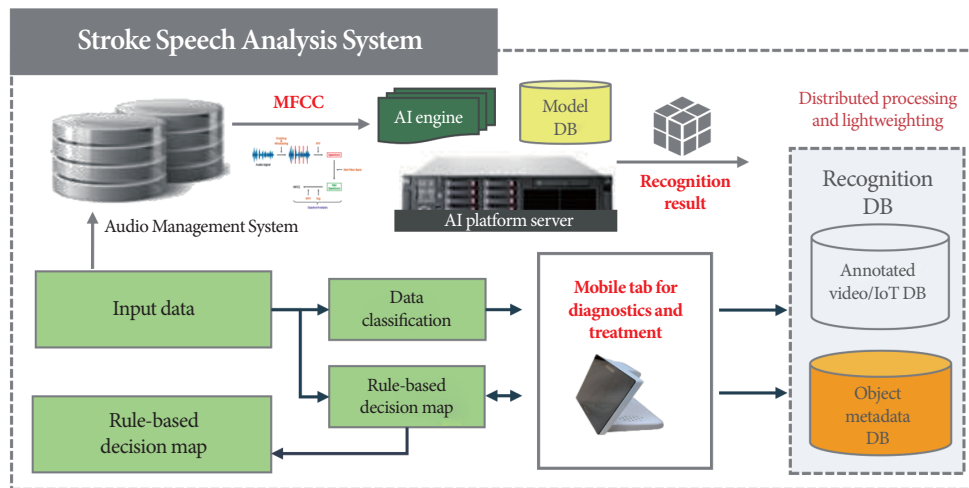

Fig. 2. Stroke speech analysis system diagram. AI, artificial intelligence; MFCC, Mel-frequency cepstral coefficient; IoT, Internet of things; DB, database.

pronunciation.

To capture the clinical patients' voices in real time, a specific sentence format was presented through an audio microphone. The patients then vocalized the sentence, and the recording was saved in an audio data format. The patients' audio data are stored in the original sound format, rather than being compressed, to minimize any loss of quality. Out of 300 clinical patients' spoken data, a dataset consisting of 200 clinical cases who pronounced a particular type of sentence from beginning to end was filtered and utilized for learning and evaluation. The training data were labeled using these 200 patients' audio data and classified as abnormal data.

We also added another 200 control cases to the dataset, so that the total training dataset consisted of 400 cases. We used the same process to build audio data for the control group as for the patient group to build training data for feature extraction and analysis algorithms.

Development of Stroke Speech Analysis System

In order to assess the risk of stroke associated with neurogenic bladder, we developed an AI model that trains, learns, and evaluates the speech of stroke patients. This model can be employed for the diagnosis and treatment of stroke-related speech and language disorders, and it is connected to a dedicated mobile tablet. The system processes and categorizes abnormalities in speech data entered through the tablet, generating alarm events as needed. The identified results are compiled into a separate recognition database, which can be used for distributed pro-
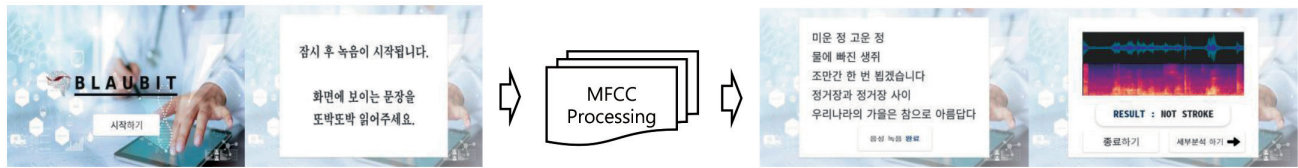
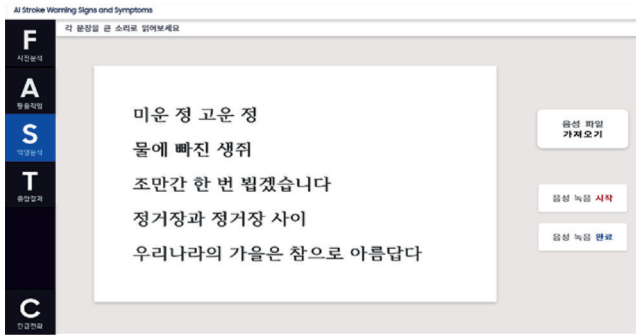**Fig. 3.** The concept process of judging speech disorders. MFCC, Mel-frequency cepstral coefficient.

Fig. 4. Main screen of the developed stroke speech analysis software.

cessing and reducing the weight of the final AI engine, thereby facilitating retraining. Fig. 2 presents a block diagram of this stroke speech analysis system.

A guideline for analysis was initially developed by assigning weights to various factors of a patient's abnormal symptoms in order to assess stroke severity. This was carried out by utilizing existing research on stroke diagnosis and modifying these guidelines to establish weighting criteria specifically for speech analysis. In this study, a learning model was constructed to identify individual stroke disorders, and accuracy was ensured in determining the presence of a stroke even with individual symptoms by utilizing a clinical patient's speech learning database. Fig. 3 below illustrates the process of evaluating speech disorders.

The first step, based on the input data, involves classifying the abnormal area using a convolutional neural network (CNN) algorithm [12-17]. Additionally, the Mel-frequency cepstral coefficient (MFCC) algorithm was employed for further refinement, enabling personalized lesion tracking. The acquired data were incorporated into a database [18-21] that contained the patients' spoken data and normal individuals' spoken data. This approach facilitated the expansion of the learning data and enhancement of the algorithm.

Building a Comprehensive Patient Management Platform
The proposed speech analysis-based feature for stroke diagnosis

aims to prevent neurogenic bladder. Patient feedback is a crucial aspect of managing urinary activity. To facilitate this, a web-based (HTML5) stroke management system was developed for systematic and comprehensive patient management. The system was constructed in 3 phases: database development, back-end development, and client development. The client was developed using the Angular JS framework to optimize the speed of switching between utilized programs, while the backend was configured with Firebase. Database development employed a NoSQL-based documentation method. In conjunction with the AI diagnostic system, a feedback function was incorporated to reflect the results, categorize and store the spoken data, and enhance the algorithm. Based on the input speech data, the AI diagnostic system uses feature extraction and pattern analysis to determine the risk of eventual stroke. The risk output are evaluated to assess the accuracy of the results. Furthermore, the AI diagnosis system was designed as a platform to allow individual disease history inquiries and support API utilization through integration with the hospital's electronic medical record system. This linkage allows for efficient clinician support and performance evaluation.

RESULTS

The performance evaluation environment proposed in this paper consists of the following elements. We have developed an AI software that collects specific speech data from clinical stroke patients, analyzes it, and provides early warnings for stroke-related neurogenic bladder issues. This software extracts feature values by analyzing the input speech, converts them into a stroke index, and stores the data. The stored stroke index is then categorized into learned stroke indicators and ultimately displayed to indicate whether a stroke has occurred. Fig. 4 illustrates the main screen of the developed stroke speech analysis software, while Table 1 presents the primary environmental information of the developed system.

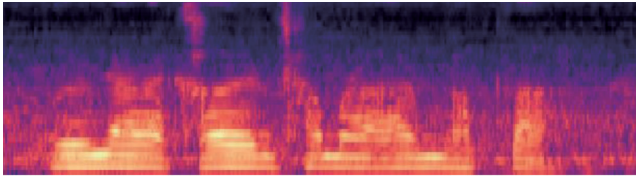The performance evaluation process was as follows: First, to

Fig. 5. Main feature spectrum (stroke speech) of Mel-frequency cepstral coefficient algorithm.

Table 1. Configure the environment of the developed system

Python	Python 3.7.6
Tensorflow	tensorflow 2.4.0, tensorflow-gpu 2.4.0, tensorboard 2.6.0
Keras	keras 2.6.0, keras-preprocessing 1.1.2
OpenCV	oencv-python 4.5.3.56, opencv-contrib-python 4.5.3.56
Mediapipe	mdiapipe 0.8.7.1
Librosa	librosa 0.8.1

analyze speech disorders caused by facial paralysis, a precursor symptom of stroke, based on AI, spoken audio data of a specific sentence from a clinical patient were acquired. The obtained spoken audio data were stored in an audio format with minimal compression loss. For performance evaluation, data acquisition and processing were performed first, consisting of data input and analysis, result display, and database transfer. The stored spoken data were feature-vectorized using the MFCC algorithm. Additionally, to build a comparison group database, the same specific sentence was presented to a healthy individual to obtain their spoken data. The same preprocessing process as the clinical patient was performed to vectorize the features. To learn the difference between the extracted feature data of the 2 groups and distinguish between patients and healthy individuals in real-time audio data, it is necessary to extract features from the spoken audio databases of the 2 groups and learn the difference. The first input speech was processed for feature analysis by the MFCC algorithm, and the CNN learning algorithm was used to extract the speech disorder index and derive classification values by index. The MFCC algorithm was employed to extract feature data from the speech to learn the differences between patients and healthy individuals. Fig. 5 illustrates the main feature spectrum of the MFCC algorithm, which is the result of the feature vectorization process based on the input speech data. Finally, through the detection of spoken stroke symptoms, at-risk individuals can be identified in advance and notification services can be provided. The user's results are stored in the database and utilized for retraining. Fig. 6 displays a data process-
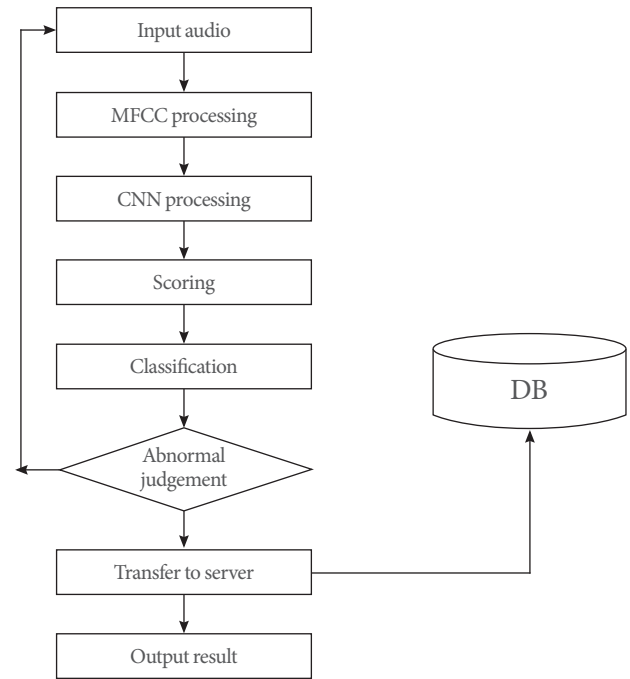

Fig. 6. The data processing flow chart for performance evaluation. MFCC, Mel-frequency cepstral coefficient; CNN, convolutional neural network; DB, databese.

ing flowchart for performance evaluation.

To assess the performance of the software, we primarily focused on the validation accuracy and training accuracy of the data. Next, we applied the analysis model by inputting both abnormal and normal data and examined the outcomes. First, we evaluated the results of the training data. A total of 400 data points, consisting of 200 abnormal and 200 normal data points, were trained with 300 epochs and a batch size of five. Fig. 7 displays a graph of the training data results, while Table 2 presents the numerical results. Subsequently, we tested the analysis model by processing 30 abnormal and 30 normal data points in real time. The results yielded 98.7% accuracy for normal data and 99.6% accuracy for abnormal data, indicating a high test accuracy. Table 3 presents the final test results.

DISCUSSION

Patients with stroke-induced neurogenic bladder who arrive at the hospital within the "golden" time frame of 6 hours after stroke and receive appropriate treatment may still face significant physical and cognitive impairments, as well as requiring extensive rehabilitation. Moreover, many stroke survivors who
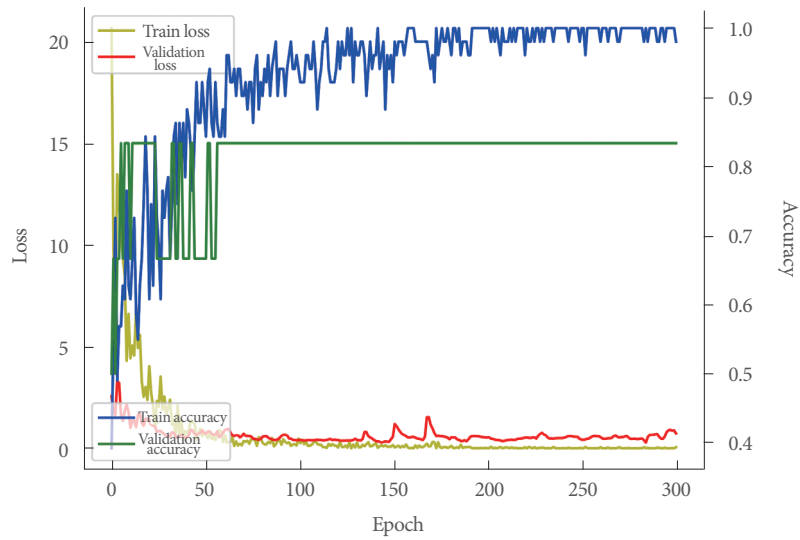
Fig. 7. Result analysis of test data (train accuracy/loss, validation accuracy/loss).

Table 2. Training result

Variable	Result
Validation accuracy	0.83
Train accuracy	0.99

Table 3. Test result

Validation accuracy	Normal	Abnormal
Dataset	30	30
Accuracy	98.7%	99.6%

do not receive treatment within this "golden hour" often experience more severe and diverse consequences than those treated earlier, leading to a diminished quality of life for both the individual and their family. Consequently, stroke is regarded as a critical condition that necessitates risk management, and the demand for incorporating various information technologies for this purpose is progressively growing.

In this study, we developed a system that delivers a final alert by analyzing speech to detect the early risk of stroke. The accuracy of the speech analysis was enhanced by employing deep learning technology based on the MFCC method, which was specifically designed for signal data analysis. To further improve accuracy, we constructed speech learning data in accordance with stroke diagnosis guidelines and designed an analysis model based on the resulting data. The factors utilized in the design were developed using speech-related information as input values among the information used to assess stroke severity. To

evaluate the accuracy of the developed software, we created a dataset with 400 training data points and 60 evaluation data points. Consequently, we achieved relatively high accuracy rates of 98.7% for normal data and 99.6% for abnormal data.

In future research, we aim to transform the software into a global stroke alert service platform by tailoring it to different countries and obtaining evaluation databases for various nations following performance assessments. In today's ultra-aging society, the prevalence of chronic diseases is on the rise, making it both necessary and significant to implement digital treatments for conditions that lead to severe complications, such as stroke. Furthermore, this AI-driven healthcare convergence medical device is anticipated to reduce national social costs by allowing patients to visit hospitals within the critical "golden hour" and receive safe treatment through mobile services and cloud servers, accessible anytime and anywhere.

AUTHOR CONTRIBUTION STATEMENT

· Conceptualization: *KJC*
· Data curation: *DJS*
· Formal analysis: *ESK*
· Funding acquisition: *ESK*
· Methodology: *KJC*
· Project administration: *KJC*
· Visualization: *ESK*
· Writing - original draft: *ESK*
· Writing - review & editing: *STC*

ORCID

Eui-Sun Kim   0000-0002-2415-9433

Dong Jin Shin   0000-0002-4390-2843

Sung Tae Cho   0000-0002-4691-6159

Kyung Jin Chung   0000-0003-0153-7814

REFERENCES

1. Özcan F, Özişler Z. The relationship between urinary symptom severity and functional status in patients with stroke. Scott Med J 2022;67:64-70.

2. Musco S, Giraudo D, Antoniono E, Lombardi G, Del Popolo G, Li Marzi V, et al. Prevalence of nocturia after brain injury: a cross-sectional study in a single rehabilitation center. Brain Inj 2021;35:90-5.

3. Kim JY, Kang K, Kang J, Koo J, Kim DH, Kim BJ, et al. Executive summary of stroke statistics in Korea 2018: a report from the epidemiology research council of the Korean Stroke Society. J Stroke 2019;21:42-59.

4. Kim YS, Park SS, Bae HJ, Cho AH, Cho YJ, Han MK, et al. Stroke awareness decreases prehospital delay after acute ischemic stroke in Korea. BMC Neurol 2011;11:2.

5. Hong KS, Bang OY, Kim JS, Heo JH, Yu KH, Bae HJ, et al. Stroke statistics in Korea: part II stroke awareness and acute stroke care, a report from the korean stroke society and clinical research center for stroke. J Stroke 2013;15:67-77.

6. Park TH, Ko Y, Lee SJ, Lee KB, Lee J, Han MK, et al. Identifying target risk factors using population attributable risks of ischemic stroke by age and sex. J Stroke 2015;17:302-11.

7. Kim JW, Lee KJ, Yang HR, Chang JY, Moon JS, Khang YH, et al. Prevalence and risk factors of elevated alanine aminotransferase among Korean adolescents: 2001-2014. BMC Public Health 2018; 18:617.

8. Chae J, Seo MY, Kim SH, Park MJ. Trends and risk factors of metabolic syndrome among Korean adolescents, 2007 to 2018. Diabetes Metab J 2021;45:880-9.

9. Hong KS, Bang OY, Kang DW, Yu KH, Bae HJ, Lee JS, et al. Stroke statistics in Korea: part I. Epidemiology and risk factors: a report from the korean stroke society and clinical research center for stroke. J Stroke 2013;15:2-20.

10. Kamal N, Sheng S, Xian Y, Matsouaka R, Hill MD, Bhatt DL, et al. Delays in door-to-needle times and their impact on treatment time and outcomes in get with the guidelines-stroke. Stroke 2017;48: 946-54.

11. Gaidhani BR, Rajamenakshi RR, Sonavane S, Brain stroke detection using convolutional neural network and deep learning models. In: 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT); Jaipur, India; 2019 Sep. 28–29. doi: 10.1109/ICCT46177.2019.8969052.

12. Kim ES, Heo JM, Eun SJ, Lee JY. Development of early-stage stroke diagnosis system for the elderly neurogenic bladder prevention. Int Neurourol J 2022;26(Suppl 1):S76-82.

13. Chin CL, Lin BJ, Wu GR, Weng TC, Yang CS, Su RC, et al. An automated early ischemic stroke detection system using CNN deep learning algorithm. In: 2017 IEEE 8th International conference on awareness science and technology (iCAST); Taichung, Taiwan; 2017 Nov. 8-10. 2017;368-72.

14. Elbagoury BM, Vladareanu L, Vlădăreanu V, Salem AB, Travediu AM, Roushdy MI. A hybrid stacked CNN and residual feedback GMDH-LSTM deep learning model for stroke prediction applied on mobile AI smart hospital platform. Sensors (Basel) 2023;23:3500.

15. Chavva IR, Crawford AL, Mazurek MH, Yuen MM, Prabhat AM, Payabvash S, et al. Deep learning applications for acute stroke management. Ann Neurol 2022;92:574-87.

16. Fang G, Huang Z, Wang Z. Predicting ischemic stroke outcome using deep learning approaches. Front Genet 2022;12:827522.

17. Bijalwan V, Semwal VB, Singh G, Mandal TK. HDL-PSR: modelling spatio-temporal features using hybrid deep learning approach for post-stroke rehabilitation. Neural Process Lett 2023;55:279-98.

18. Wang H, Wu Z, Ma S, Lu S, Zhang H, Ding G, et al. Deep learning for signal demodulation in physical layer wireless communications: prototype platform, open dataset, and analytics. IEEE Access 2019; 7:30792-801.

19. Wang Y, Liu M, Yang J, Gui G. Data-driven deep learning for automatic modulation recognition in cognitive radios. IEEE Trans Veh Technol 2019;68:4074-7.

20. Chen W, Lei X, Chakrabortty R, Chandra Pal S, Sahana M, Janizadeh S. Evaluation of different boosting ensemble machine learning models and novel deep learning and boosting framework for head-cut gully erosion susceptibility. J Environ Manage 2021;284:112015.

21. Subramaniyam M, Lee KS, Park SJ, Min SN. Development of mobile application program for stroke prediction using machine learning with voice onset time data. In: Stephanidis C, Antona M, editors. HCI International 2020 - Posters. HCII 2020. Communications in Computer and Information Science, vol 1224. Cham: Springer; 2020:1224:670-5. https://doi.org/10.1007/978-3-030-50726-8_87.