Technology of Neurological Disease Recognition Using Gated Recurrent Unit Neural Network and Internet of Things

Uladzimir Vishniakou and Yiwei Xia and Chuyue Yu
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
Email: vish@bsuir.by, minskxiayiwei@qq.com, ycy18779415340@gmail.com

***Abstract*—In this paper, authors proposed a neurological disease recognition technique using gated recurrent unit neural network and supporting Internet of Things (IoT), which was checked by taking Alzheimer's disease (AD) and Parkinson's disease (PD) as examples. In this method first pre-emphasized and denoised the voice data, then segmented the voice signals with a sliding fixed window using the Hamming window function. Then we were extracted the eGeMAPSv02 voice features from the window signal, fed the features into the gated recurrent unit neural network model for its training, testing and achieve the disease diagnosis. The results of the study showed that despite the limited generalization ability of the gated recurrent unit model, it can still efficiently achieve voice recognition detection of a portion of neurological diseases. The model is implemented on the basis of the IoT platform for building a subsystem of IT diagnostics of patients as part of the smart city project. The code is stored in https://github.com/HkThinker/Technology-of-neural-disease-recognition.**

***Keywords*—gated recurrent unite neural network, Internet of things, voice recognition, neurological disease**

I. INTRODUCTION

Neurological disease usually result in structural or functional changes in the nervous system, causing patients to suffer from perception, thinking, emotion and behavior, and present a significant challenge to the global healthcare system. They are a group of diseases that affect the nervous system and include a variety of disorders such as neurodegenerative diseases, autoimmune diseases, cerebrovascular diseases, and brain injuries. For example, PD is a neurodegenerative disease that affects motion management and is characterized by symptoms such as hand tremors, limb stiffness, slow movements, and postural instability. AD is similar and results in memory loss, cognitive decline, and abnormal language and behavior. They tend to occur in older age groups, currently have no complete medical cure, but early diagnosis and prompt treatment can alleviate symptoms and slow progression. Traditional diagnosis of neurological diseases is usually based on doctors' clinical experience, medical history, physical examination and specific tests, which has limitations and requires a lot of labor and resources. In recent years, with the rapid development of artificial intelligence and IoT technologies, neurological disease identification technologies using neural networks and supporting IoT are expected to become a new breakthrough point.

The main purpose of this paper is to investigate the Gated Recurrent Unit (GRU) neural networks and IoT technologies to recognition for neurological diseases. To be specific, our research aims to achieve the following objectives:

1) To develop a GRU neural network model, which was trained through a publicly available database to implement the diagnosis and prediction of PD and AD.
2) By using IoT technology, we collected patients' voice data and combined these data with the GRU neural network model to improve the precision and accuracy of diagnosis and prediction of neurological diseases.
3) To deploy the GRU neural network model to the Thingspeak IoT platform.

II. RELATED WORK

A. Application of IoT in Neurological Disease Diagnosis

Neurological disease diagnosis systems that are based on neural network technology and IoT technology have been widely used.

B. Lu [1] built a practical brain MRI-based AD diagnostic classifier using deep learning/transfer learning on datasets of unprecedented size and diversity. The purpose of Mukherji [2] was to identify non-invasive, inexpensive markers and develop neural network models that learn the relationship between those markers and the future cognitive state. David Payares-Garcia [3] proposed a classification technique that incorporates uncertainty and spatial information for distinguishing between healthy subjects and patients from four distinct neurodegenerative diseases: AD, mild cognitive impairment, PD, and Multiple Sclerosis. Abbas Sheikhtaheri [4] aimed to identify and classify the IoT technologies used for AD dementia as well as the healthcare aspects addressed by these technologies and the outcomes of the IoT interventions.

Researchers had identified the feasibility of integrating deep learning, cloud, and IoT, Syed Saba Raoof [5] explained a summary of various techniques utilized in smart healthcare, i.e., deep learning, cloud-based-IoT applications in smart healthcare, fog computing in smart healthcare, and challenges and issues faced by smart healthcare and it presents a wider scope as it is not intended for a particular application such aspatient monitoring, disease detection, and diagnosing and the technologies used for developing this smarta systems are outlined. Reyazur Rashid Irshad [6] proposed a novel healthcare monitoring system that tracks disease processes and forecasts diseases based on the available data obtained from patients in distant communities. Rafael A Bernardes [7] presented a perspective on integrating wearable technology and IoT to support telemonitoring and self-management of people living with PD in their daily living environment.

B. Classification of Voice Features

Since more than 90 % of PD patients have varying degrees of dysphonia in the early stages of the disease, the diagnosis of PD based on voice features has the merits of being non-invasive and convenient. Darley [8] first used voice to diagnose aphasia in 1969. Saker et al. [9] preprocessed the voice data and extracted features, then applied SVM and KNN classification algorithms to the feature matrix for classification, eventually obtaining an average accuracy and a maximum accuracy of 55 % and 85 %, respectively, which initially confirmed the feasibility of voice features to classify PD. To further improve the accuracy of model prediction and simplify the algorithmic model, scholars have applied different feature selection algorithms.

III. Methodology and Datasets

A. Pre-emphasis and Denoising of Voice Signals

It is difficult to obtain the high-frequency part of the unprocessed voice signal because the power of the voice signal will be significantly attenuated after the sound gate excitation as well as the influence of mouth and nose radiation, combined with the smaller energy corresponding to the high frequency while the larger energy corresponding to the low frequency in the spectrogram of the voice signal. In order to facilitate the spectrum analysis, this paper adopted a first-order FIR high-pass digital filter for the pre-emphasis processing of the voice signal. The purpose of pre-emphasis is to improve the high-frequency part, so that the spectrum of the voice signal becomes flat, thus the spectrum can be obtained with the same signal-to-noise ratio in the whole frequency band.

Voice denoising is an effective part of signal preprocessing, mainly to improve the quality of voice and obtain more pure voice signals. The Fig. 1 shows the process of voice signal denoising.

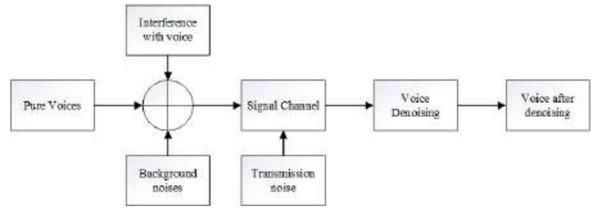

Figure 1. Flow chart of voice denoising.

According to the different parts of the noise introduction, voice noise can be divided into background noise and transmission noise. In this paper, spectral subtraction algorithm was used to denoise the voice. The spectral subtraction algorithm is designed based on the principle that pure voice is statistically independent of the noise signals.

B. Framing and Windowing of Voice Signals

The Fourier transform commonly used in voice signal processing calls for a smooth signal, but the main feature of the voice signal is the short-time smoothness, i.e., the stability of the voice signal in 10–30 ms period. Therefore, if we want to characterize the voice signal, it is necessary to analyze the short-time characteristics of the voice signal, the original signal is framed, and the frame frequency signal with short-time smoothness is derived. In the process of frame splitting, the signal tends to produce spectral deficiencies, so a windowing process must be performed between frames to keep the signal at the truncation without distortion. The windowing function used in this paper is the Hamming window function, with window size of 1024, frequency of voice signal is 44.1khz, and the overlap rate of window is 50 %, hence the voice time of one window is about 23 ms.

C. Feature Extraction of Voice Signals

We used an extended version of GeMAPS (Basic Affective Parameter Set), eGeMAPSv02 [10], a speech feature set. It uses acoustic features and spectral-based features to describe the speech signal, with a total of 88 features. It contains 25 low-level descriptor features, namely pitch, jitter, gating frequency, gating bandwidth, gloss, loudness, harmonic-to-noise ratio (HNR), Alpha ratio, Hammarberg index, spectral slope 0–500 Hz, spectral slope 500–1500 Hz, 3 gating relative energies, 3 relative energies, 3 harmonic differences, 4 Mel–Frequency Cepstral Coefficients, 1 spectral flux. 53 other parameters are derived from these basic parameters.

D. 6–layer Gated Recurrent Unit Model

In the paper, a multi-layer GRU model is constructed for voice data recognition. two mechanisms, an update gate and a reset gate, are included in the GRU module. The internal equation of a single GRU model is :

$$r_t = \sigma(W_r \times [h_{t-1}, x_t]) \quad (1)$$

$$z_t = \sigma(W_z \times [h_{t-1}, x_t]) \quad (2)$$

$$\widetilde{h}_t = tanh(W_{\widetilde{h}} \times [r_t \times h_{t-1}, x_t]) \quad (3)$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \widetilde{h}_t \quad (4)$$

$$y_t = \sigma(W_\sigma \times h_t) \quad (5)$$

Where σ represents sigmod function, z_t is the update gate of the unit, sigmod function converges the value of the update gate to 0 or 1, i.e., whether the value of the previous step is remembered or discarded. r_t is the reset gate, the smaller r_t, the more information about the previous state needs to be ignored, W is the weight value, h_t and $\widetilde{h}_t$ are the output and temporary hidden states in the module.

The GRU model has a lower computational cost with faster training, so the model is extensively used in various fields of deep learning. The structure of a single GRU module is shown in Fig. 2.

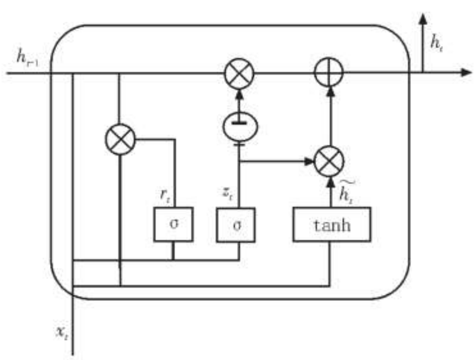

Figure 2. Structure of a single GRU model.

The GRU neural network in this work had a total of 6 layers and 1700 learnable properties. Table III illustrated the structure of the GRU neural network.

Table I
THE STRUCTURE OF GRU NEURAL NETWORKS

Name	Activations	Learnable Properties
Sequence Input	$88(C) \times 1(B) \times 1(T)$	
GRU	$6(C) \times 1(B)$	InputWeights 18×88 RecurrentWeights 18×6 Bias 18×1
ReLU	$6(C) \times 1(B)$	
Fully Connected	$2(C) \times 1(B)$	Weights 2×6 Bias 2×1
Softmax	$2(C) \times 1(B)$	
Classification Output	$2(C) \times 1(B)$	

E. Public Voice Datasets Used

This paper used public datasets [11] collected from 188 PD patients (107 men, 81 women) aged 33–87 at Istanbul University. The control group includes 64 healthy individuals (23 men, 41 women) aged 41–82. Participants were asked to sustainably pronounce the vowel /a/ while a microphone set at 44.1 KHz recorded their voice three times.

DementiaBank [12] is a resource that collects voice, video, and text data from older adults and patients with AD. It contains two groups of participants; the elderly group includes 60 healthy older adults from New York City who ranged in age from 60 to 91 years, while the AD group includes 64 patients from Pittsburgh who ranged in age from 60 to 95 years. Each participant was asked to answer a series of questions. Data were collected using a specialized recording device, with recorded voice data at a sampling rate of 44.1 kHz.

Visualization of voice can help to extract feature information. The voice waveform and spectrum of AD are shown in Fig. 3.

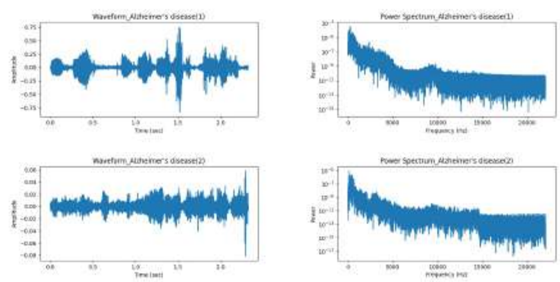

Figure 3. Alzheimer data voice waveforms and spectrograms.

F. Training and Testing Process

After noise removal and signal segmentation of all the voice data in the dataset, we extracted 88 voice signal features for each voice window signal, and then created a neural network for training and learning. In this paper, a 6-layer GRU model was adopted. It's model structure is shown in Fig. 4.

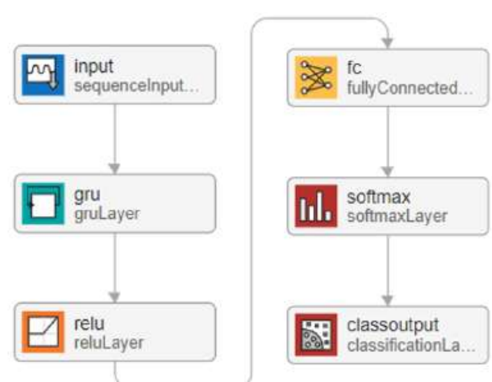

Figure 4. Structure of 6-layer GRU model.

To avoid overfitting, we added a Relu layer after the GRU layer and output the probability of class labels by defining a Softmax layer to vectorize the labels onte-hot before calculating the correctness. To accelerate the model convergence, the training took a batch gradient descent approach for weight update, and each batch contained 64 features. A flow chart of the overall model structure of the experiment is shown in Fig. 5.

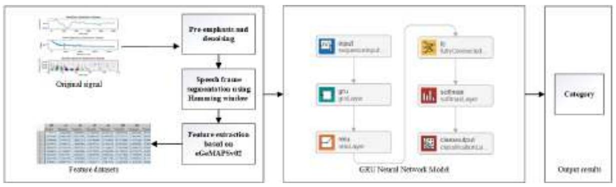

Figure 5. Flow chart of the overall model structure of the experiment

G. Deploying the Model to the Thingspeak Platform

Thingspeak is an open source IoT application platform that allows users to conveniently collect, process and analyze data from IoT devices. The platform provides a way for developers and manufacturers to collect, store, analyze and visualize data from the center of IoT devices and use that data for real-time decision-making and operations.

To upload the sensor data from the mobile phone to the Thingspeak IoT platform and read the results of the data analysis using an application developed by ourselves, we followed the following steps:

1) Registering a Thingspeak account and creating new channel.
2) Getting our channel write/read API Key, which we can find in our Thingspeak account.
3) Adding network authority and sensor permission in our application.
4) Adding the Thingspeak API library, we can get the source code of the library from the Thingspeak website.
5) Implementing the code to upload data in our application, the code should use HTTP protocol to upload our sensor data to our Thingspeak channel, providing the channel write API Key to authenticate our identity.
6) After uploaded data, we can analyze the data using Thingspeak's analytics tool. Once we have uploaded the data, we could use Thingspeak's analytics tool to analyze the data. To get the result of data analysis by using HTTP GET request.
7) Implementing the code to read the analysis results in our application. We need to get data analysis results using HTTP GET request and read API key to parse the results into JSON format so that we can process and display the data in our application.

In summary, to upload the sensor data from the phone to the Thingspeak IoT platform and read the data analysis results, we need to register an account and create a new channel, get the channel write/read API Key, add network permissions and sensor permissions, add the Thingspeak API library, implement the code to upload the data, use the platform's analysis tool to analyze the data, implement the code to read the analysis results, and parse the results into JSON format to process as well as display the data in the application.

Deployment of the GRU model to the Thingspeak IoT platform for data analysis Data analysis on Thingspeak using the GRU model involves the following steps:

1) Creating a new channel on Thingspeak to store the data to be analyzed. We can use Thingspeak's REST API or MQTT API to add the sensor data to the channel.
2) Training a GRU model on our local computer and exporting the model to a format that can be used on.
3) Uploading the exported KNN model to the Thingspeak platform. We can use Thingspeak's REST API or MQTT API to upload the model to the channel.
4) Once the model is uploaded successfully, we can use Thingspeak's MATLAB analysis toolbox or matlab scripts to load the model and classify the uploaded data. In MATLAB, we can read the uploaded data using the thingSpeakRead function, load the GRU model using the load function, and classify the data using the predict function.
5) Displaying the classification results on the user interface of Thingspeak or sending the results to our cell phone as well as to an email for easy viewing of the identification results.

IV. Experiments and Results

A. Experimental Setup

In this paper, the feature datasets were divided into training datasets and test datasets in the ratio of 9:1. The training datasets were trained and validated using the 5-fold cross-validation method, which was repeated five times. The test datasets were then used to test the final results. And We evaluated the experiment using the confusion matrix [13].

The Table II showed the GRU neural network model hyperparameter setting table in this experiment.

B. Experiment Results and Evaluation

The Fig. 6 showed the process of training the GRU model in 1000 epochs based on the Parkinson's public voice datasets.

As seen in the Fig. 6, the GRU neural network model based on the Parkinson's public voice dataset can converge substantially in a short time. The model uses stochastic gradient descent and variable learning rate in solving the minimization loss function, so there was

Table II
GRU NEURAL NETWORK MODEL HYPERPARAMETER SETTING

Number	Parameter Name	Parameter Value
1	Mini Batch Size	64
2	Max Epochs	1000
3	Initial Learn Rate	0.01
4	Learn Rate Drop Factor	0.1
5	Learn Rate Drop Period	700
6	Shuffle	every-epoch
7	optimization	adam

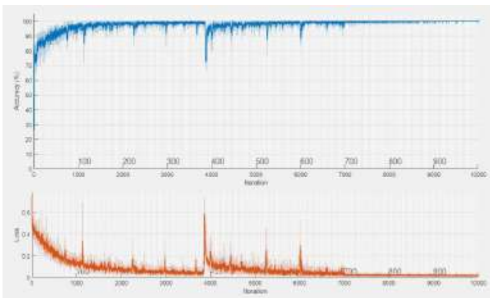

Figure 6. The process of training the GRU model based on Parkinson's datasets in 1000 epochs.

some jitter in the convergence process of the model, but the general trend of the model accuracy was improved, the loss function of the model corresponded to a decreasing trend. The final training accuracy of the model reached 100 %.

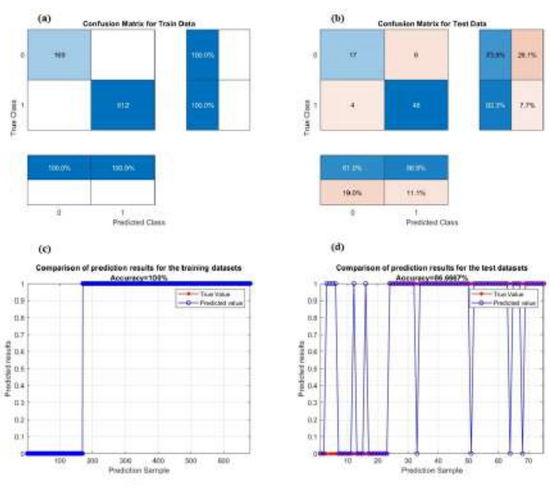

Figure 7. a - confusion matrix of training datasets; b - confusion matrix of testing datasets; c - prediction results for the training datasets; d - prediction results for the testing datasets.

As can be seen from Fig. 7, the accuracy of the model on the test set was 86.66 %, while the accuracy on the training set was much better than the accuracy on the test set, the model may have been overfitted. The overfitting phenomenon may arise because of the small amount of data in the public voice dataset of Parkinson's, coupled with the uneven distribution of samples in this public dataset, so the model's performance was degraded.

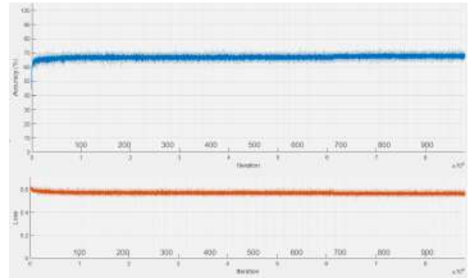

Figure 8. The process of training the GRU model based on Alzheimer's datasets in 1000 epochs.

As can be seen in Fig. 8, the model converged after the Alzheimer's voice training dataset was fed into the model and entered 2000 training cycles. The Fig. 9 showed a comparison of the prediction results and the confusion matrix of the training and testing datasets for AD.

Figure 9. a - confusion matrix of testing datasets; b - confusion matrix of training datasets.

Table III showed the experimental results of PD recognition and AD using GRU based on the test datasets.

Table III
THE EXPERIMENTAL RESULTS OF PD RECOGNITION AND AD USING GRU BASED ON THE TEST DATASETS

Public Datasets	Average Precision	Average Sensitivity	Average F1 score	Test Accuracy
Parkinson's	84.95 %	83.10 %	84.01 %	86.66 %
Alzheimer's	67.60 %	67.50 %	67.55 %	68.27 %

In summary, the accuracy of the GRU-based PD model could reach 86.67 % on the test dataset and 100 % on the training dataset. On the testing datasets, the average precision was 84.95 %, the average sensitivity was 83.10 %, and the average F1 score was 84.01 %. This experimental result showed that the recognition of PD using GRU algorithm based on freezing of gait data was effective.

However, the test results of the model on Alzheimer's data were not satisfactory, which may be due to the fact that Alzheimer's data were more complex and harder to find feature points compared to Parkinson's data, after which we will try new models or improve the model in a way to increase the accuracy. The model is implemented on the basis of the IoT platform for building a subsystem of IT diagnostics of patients as part of the smart city project using elements of OSTIS technology [14].

V. Conclusion

The aim of this paper was to explore the performance of GRU neural networks in voice recognition tasks within neurological diseases. We used a 6-layer GRU model that was trained and tested on the Parkinson's public voice dataset and the Alzheimer's public voice dataset. With the experimental results, we found that.

1) On the Parkinson's public voice dataset, our model can achieve 86.66 % accuracy, which has better performance than traditional machine learning methods. However, the accuracy on the Alzheimer public voice dataset was only 68.27 %, indicating that the 6-layer GRU model does not have good generalization ability.
2) During the training of the model, we noticed that the training error of the model was gradually decreasing with the increase of training times, but the testing error started to increase. This indicated that the model appeared overfitting phenomenon.
3) We also explored our scheme to implement the GRU model on the IoT. The scheme has potential for practical applications and provides a reference for research in related fields.

Summing up, our experimental results showed that the GRU model can be deployed on IoT platforms to solve part of the problem of IT diagnostics of neurological disorders by recognizing changes in patients' speech.

References

[1] B. Lu, B. Li, H. Chang, Z. Li, L. Chen, N. Zhu, Z. Zhou, H. Li, X. Wang, Y. Cui *et al.*, "A practical alzheimer disease classifier via brain imaging-based deep learning on 85,721 samples [preprint]," *Neuroscience*, 2020.

[2] D. Mukherji, M. Mukherji, and N. Mukherji, "Early detection of alzheimer's disease using neuropsychological tests: a predict–diagnose approach using neural networks," *Brain Informatics*, vol. 9, no. 1, pp. 1–26, 2022.

[3] D. Payares-Garcia, J. Mateu, and W. Schick, "Spatially informed bayesian neural network for neurodegenerative diseases classification," *Statistics in medicine*, vol. 42, no. 2, pp. 105–121, 2023.

[4] A. Sheikhtaheri and F. Sabermahani, "Applications and outcomes of internet of things for patients with alzheimer's disease/dementia: A scoping review," *BioMed Research International*, vol. 2022, 2022.

[5] S. Saba Raoof, M. Durai *et al.*, "A comprehensive review on smart health care: Applications, paradigms, and challenges with case studies," *Contrast Media & Molecular Imaging*, vol. 2022, 2022.

[6] R. R. Irshad, A. A. Alattab, O. A. S. Alsaiari, S. S. Sohail, A. Aziz, D. i. Madsen, and K. M. Alalayah, "An optimization-linked intelligent security algorithm for smart healthcare organizations," in *Healthcare*, vol. 11, no. 4. MDPI, 2023, p. 580.

[7] R. A. Bernardes, F. Ventura, H. Neves, M. I. Fernandes, and P. Sousa, "Wearable walking assistant for freezing of gait with environmental iot monitoring: A contribution to the discussion," *Frontiers in Public Health*, vol. 10, 2022.

[8] F. L. Darley, A. E. Aronson, and J. R. Brown, "Differential diagnostic patterns of dysarthria," *Journal of speech and hearing research*, vol. 12, no. 2, pp. 246–269, 1969.

[9] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a parkinson speech dataset with multiple types of sound recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828–834, 2013.

[10] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan *et al.*, "The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing," *IEEE transactions on affective computing*, vol. 7, no. 2, pp. 190–202, 2015.

[11] C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, M. Tutuncu, T. Aydin, M. E. Isenkul, and H. Apaydin, "A comparative analysis of speech signal processing algorithms for parkinson's disease classification and the use of the tunable q-factor wavelet transform," *Applied Soft Computing*, vol. 74, pp. 255–263, 2019.

[12] A. M. Lanzi, A. K. Saylor, D. Fromm, H. Liu, B. MacWhinney, and M. L. Cohen, "Dementiabank: Theoretical rationale, protocol, and illustrative analyses," *American Journal of Speech-Language Pathology*, pp. 1–13, 2023.

[13] A. Luque, A. Carrasco, A. Martín, and A. de Las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, 2019.

[14] D. Shunkevich, "Agent-oriented problem solvers of intelligent systems component," *autoref. of PhD thesis in tech. sciences: 05.13.17 / Belarusian State University of Informatics and Radioelectronics, Minsk*, pp. 1–24, 2018.

Технология распознавания нейрологических заболеваний с использованием нейронной сети закрытого рекуррентного блока и интернета вещей

Вишняков В. А., Ивей С., Чуюэ Ю.

В этой статье авторы предложили метод распознавания неврологических заболеваний с использованием закрытой рекуррентной нейронной сети и поддержки Интернета вещей (IoT), который был проверен на примере болезни Альцгеймера (БА) и болезни Паркинсона (БП). В этом методе сначала предварительно выделяются и ослабляются голосовые данные, затем голосовые сигналы сегментируются с помощью скользящего фиксированного окна с использованием функции окна Хэмминга. Затем извлекаются голосовые характеристики eGeMAPSv02 из сигнала окна, вводятся эти характеристики в модель нейронной сети с закрытым рекуррентным модулем для ее обучения, тестирования и достижения диагноза заболевания. Результаты исследования показали, что, несмотря на ограниченную способность модели gated recurrent unit к обобщению, она может эффективно обеспечивать распознавание голоса при выявлении части неврологических заболеваний. Модель реализуется на базе платформы IoT для построения подсистемы ИТ-диагностики пациентов а рамках проекта умного города. Код хранится https://github.com/HkThinker/Technology-of-neural-disease-recognition.

Ключевые слова — нейронная сеть с закрытым рекуррентным модулем, сеть Интернета вещей, распознавание голоса, неврологические заболевания.

Термины индекса — закрытая рекуррентная нейронная сеть, Интернет вещей, распознавание голоса, неврологические заболевания.

Received 12.03.2023