https://doi.org/10.51885/3134-8025_JICS_2026_2_1
SRSTI 28.23.31

COMPARATIVE EVALUATION OF CLASSIC AND DEEP LEARNING MODELS FOR EARLY DETECTION OF DYSARTHERIA

ДИЗАРТЕРИЯНЫ ЕРТЕ АНЫҚТАУ ҮШІН КЛАССИКАЛЫҚ ЖӘНЕ ТЕРЕҢ ОҚЫТУ МОДЕЛЬДЕРІН САЛЫСТЫРМАЛЫ БАҒАЛАУ

СРАВНИТЕЛЬНАЯ ОЦЕНКА КЛАССИЧЕСКИХ И ГЛУБОКИХ МОДЕЛЕЙ ОБУЧЕНИЯ ДЛЯ РАННЕГО ВЫЯВЛЕНИЯ ДИЗАРТЕРИИ

A.B. Aben [1], **O.Zh. Mamyrbayev** [2], **G.N. Kazbekova** [1], **N.M. Zhunissov** [1], **A.S. Baimakhanova** [1*]

[1]Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkestan, Kazakhstan
[2]Institute of Information and Computational Technologies, Almaty, Kazakhstan
*Corresponding author: Baimakhanova Aigerim Sattarovna, e-mail: aygerim.baymakhanova@ayu.edu.kz

keywords:
Dysarthria, speech disorders, speech signal processing, machine learning, deep learning, UA-Speech, TORGO, RandomForest, XGBoost, SVM, Logistic Regression, MLPClassifier, CNN, neural networks, speech diagnostics.

ABSTRACT

This article analyzes the scientific foundations and practical effectiveness of applying machine learning and deep learning methods for the automatic detection and severity assessment of dysarthria. Dysarthria is a common symptom of neurological disorders such as Parkinson's disease, cerebrovascular accidents, cerebral palsy, and multiple sclerosis, leading to impairments in articulatory, prosodic, and resonant components of speech. Due to the subjectivity and time consumption of traditional clinical diagnostics, the demand for automated speech processing systems is increasing. In this study, the UA-Speech and TORGO datasets were employed, and classical machine learning models such as RandomForest, XGBoost, LightGBM, SVM, LogisticRegression, and MLPClassifier were compared with deep learning approaches based on CNN architectures. The models were optimized using GridSearchCV, while the pronounced data imbalance was addressed with ADASYN and class weight techniques. Key acoustic features such as MFCC, Mel spectrograms, and pitch were used to construct the feature space. The results indicate that machine learning models can achieve high accuracy in detecting dysarthria symptoms, while hybrid approaches combining CNNs provide improved overall performance. This research lays the groundwork for developing automated early dysarthria diagnostic systems and contributes to the advancement of technologies aimed at clinical practice.

Түйінді сөздер:
Дизартрия, сөйлеу бұзылыстары, сөйлеу сигналдарын өңдеу, машинамен оқыту,

ТҮЙІНДЕМЕ

Бұл мақалада дизартерияны автоматты түрде анықтау және оның деңгейін бағалау мақсатында машиналық оқыту және терең оқыту әдістерін қолданудың ғылыми негіздері мен практикалық тиімділігі талданады. Дизартерия Паркинсон ауруы, ми қан айналымының

терең оқыту, UA-Speech, TORGO, RandomForest, XGBoost, SVM, Logistic Regression, MLPClassifier; CNN, нейрондық желілер, сөйлеу диагностикасы.

бұзылуы, церебралдық сал ауруы және көптеген склероз сияқты неврологиялық патологиялардың кең таралған симптомы болып табылады, сөйлеудің артикуляциялық, просодиялық және резонанстық компоненттерінің бұзылуына әкеледі. Дәстүрлі клиникалық диагностика субъективтілігі мен уақыттық шығындарына байланысты, сөйлеуді автоматты өңдеу жүйелеріне сұраныс артып келеді. Зерттеу барысында UA-Speech және TORGO датасеттері қолданылып, RandomForest, XGBoost, LightGBM, SVM, LogisticRegression және MLPClassifier сияқты классикалық машиналық оқыту модельдері, сондай-ақ CNN архитектурасына негізделген терең оқыту тәсілдері салыстырылды. Модельдер GridSearchCV арқылы оңтайландырылып, деректердегі айқын имбаланс ADASYN әдісі және class weight тәсілімен өңделді. MFCC, Mel-спектограммалар және pitch сияқты негізгі акустикалық ерекшеліктер ерекшелік кеңістігін қалыптастыру үшін пайдаланылды. Алынған нәтижелер машиналық оқыту модельдерінің дизартерия белгілерін анықтауда жоғары дәлдікке қол жеткізе алатынын, ал CNN-мен біріктірілген гибридті тәсілдердің жалпы тиімділікті жақсартатынын көрсетті. Бұл зерттеу дизартерияны ерте диагностикалаудың автоматтандырылған жүйелерін әзірлеуге мүмкіндік беріп, клиникалық тәжірибеде қолдануға бағытталған технологиялардың дамуына негіз қалайды.

Ключевые слова:
Дизартрия, речевые нарушения, обработка речевых сигналов, машинное обучение, глубокое обучение, UA-Speech, TORGO, RandomForest, XGBoost, SVM, Logistic Regression, MLPClassifier, CNN, нейронные сети, диагностика речи.

АННОТАЦИЯ

В данной статье анализируются научные основы и практическая эффективность применения методов машинного обучения и глубокого обучения для автоматического выявления дизартрии и оценки её степени. Дизартрия является распространённым симптомом неврологических заболеваний, таких как болезнь Паркинсона, нарушение мозгового кровообращения, детский церебральный паралич и рассеянный склероз, что приводит к нарушениям артикуляционных, просодических и резонансных компонентов речи. Из-за субъективности и временных затрат традиционной клинической диагностики растёт спрос на автоматизированные системы обработки речи. В ходе исследования использовались датасеты UA-Speech и TORGO, сравнивались классические модели машинного обучения — RandomForest, XGBoost, LightGBM, SVM, LogisticRegression и MLPClassifier — а также методы глубокого обучения на основе архитектуры CNN. Модели оптимизировались с помощью GridSearchCV, а выраженный дисбаланс данных обрабатывался методами ADASYN и class weight. Основные акустические признаки, такие как MFCC, Mel-спектрограммы и pitch, были использованы для формирования пространства признаков. Полученные результаты показали, что модели машинного обучения могут обеспечивать высокую точность в выявлении признаков дизартрии, а гибридные подходы с использованием CNN повышают общую эффективность. Данное исследование создаёт основу для разработки автоматизированных систем ранней диагностики дизартрии и способствует развитию технологий, ориентированных на применение в клинической практике.

INTRODUCTION

Dysarthria is a speech disorder that occurs due to a violation of the neuromuscular mechanisms involved in speech control. This disorder impairs the accuracy of articulation, voice

volume, speech rhythm, and respiratory coordination. The pathology is often associated with lesions of the central or peripheral nervous system, so brain injury, stroke, neurodegenerative diseases, and developmental disorders in children are the main causes of dysarthria. A decrease in the quality of speech limits a person's ability to communicate and negatively affects social adaptation and emotional state.

Currently, the diagnosis of dysarthria is largely based on speech therapy assessment and clinical observation. However, the subjectivity, inter-rater differences, and time-consuming nature of such traditional methods indicate the need for automated approaches. In this regard, in recent years, research on the detection of dysarthria using speech signal processing, acoustic feature analysis, and machine learning technologies has gained significant momentum. Deep learning, ensemble models, and automatic spectral parameter extraction methods allow us to refine and accelerate the diagnostic process.

The study of dysarthria is important not only for clinical practice, but also for digital health, telemedicine, and technologies that use human speech as an interface. Automated systems allow us to detect speech disorders at an early stage, create individual rehabilitation programs, and dynamically monitor the patient's condition. The relevance of this research is to identify effective ways to detect dysarthria by developing highly accurate, reliable, and data-independent machine learning models.

LITERATURE REVIEW

Dysartheria is a motor speech disorder caused by motor dysfunction of the speech muscles, which disrupts elements of speech such as articulation, prosody, resonance, and breathing. It is a common symptom of neurological diseases such as Parkinson's disease (PD), cerebrovascular accident (CVA), cerebral palsy (CP), and multiple sclerosis (MS). Dysarthria impairs speech ability, significantly impairs the patient's quality of life, and leads to social isolation and psychological problems (NIDCD, n.d.).

According to statistical data, the prevalence of dysarthria varies depending on the underlying disease. Approximately 90% of people with Parkinson's disease have dysarthria, which significantly impairs their ability to speak (Jayaraman & Das, 2023).

In recent years, artificial intelligence (AI) and machine learning (ML) methods have been actively studied for the early detection and assessment of dysarthria, as traditional clinical methods are subjective and time-consuming (Prabhala et al., 2025).

Dysarthria is present in 40-51% of people with multiple sclerosis, and is associated with disease progression (Smyrni et al., 2025).

Dysarthria is the primary speech disorder in 50-90% of people with cerebral palsy (Cerebral Palsy Alliance, n.d.).

Approximately 5% of children have speech disorders by the first grade, including dysarthria, stuttering, and speech sound disorders (NIDCD, n.d.).

Dysarthria after stroke occurs in about half of patients in the UK, increasing the need for speech and language therapy (Lin et al., 2025).

The global prevalence of dysarthria is increasing due to the rise of neurodegenerative diseases, especially among the elderly (Hassan et al., 2025).

Recent studies have significantly increased the use of deep learning (DL) and machine learning (ML) models for dysarthria detection, such as CNN, LSTM, and XGBoost (Shih et al., 2022). These methods have been tested on the UA-Speech and TORGO datasets, which cover different degrees of dysarthria (Bhat & Strik, 2025).

The UA-Speech dataset consists of speech from 15 dysarthria and 13 healthy speakers, while the TORGO dataset consists of data from 8 dysarthria and 7 healthy speakers (Varma et al., 2025). The main achievements have focused on the automatic extraction of speech features, such

as MFCC, Mel-spectrograms, and pitch. However, data imbalance, noise, and a small number of speakers reduce the generalizability of the models (Wihlborg et al., 2025).

The aim of this study is to propose an effective method for dysarthria detection by comparing RandomForest, XGBoost, LightGBM, SVM, LogisticRegression, and MLPClassifier models on the UA-Speech and TORGO datasets. These models are tuned using GridSearchCV, and the imbalance is corrected using ADASYN and class weights, and are also combined with CNNs. As a result, a basis for clinical applications of early detection of dysarthria is formed, which contributes to improving the quality of life of patients (Vogel et al., 2025).

In recent years, research on dysarthria detection and severity classification has increased significantly, mainly due to the development of machine learning and deep learning methods. Datasets such as UA-Speech and TORGO are widely used, and models include methods such as CNN, LSTM, and XGBoost. Table 1 below compares recent studies, showing their methods, datasets, and results. This table highlights the main trends in the research: the dominance of deep learning models, accuracy ranging from 60-94%, and imbalance issues.

Table 1. Comparative description of models, datasets and their accuracy results used in dysarthria detection

No	Author(s) (Year)	Method	Dataset	Main Result
1	Merler et al. (2025)	Attention-based deep learning	ALS dataset	Accuracy 86.9%
2	Shahamiri et al. (2025)	DCNN with MFLS	Custom dataset	Accuracy 93.97%
3	Anuprabha et al. (2025)	Multi-modal (speech + text)	UA-Speech	Accuracy 81.8%
4	Lin et al. (2025)	Deep learning for stroke-associated	Stroke dataset	Accuracy 70%
5	Javanmardi et al. (2024)	Pretrained speech embeddings	Heterogeneous datasets	Accuracy 62.5%
6	Vogel et al. (2025)	Dysarthria Impact Scale	Custom dataset	Accuracy 79.44%
7	Stuart (2025)	Deep learning for severity	TORGO	Accuracy 75.80%
8	Alharbi et al. (2025).	Automatic classification of speech	General	Accuracy 86.9%
9	NIDCD (n.d.)	Statistics	General	Accuracy 70–80%
10	Cerebral Palsy Alliance (n.d.)	Fact sheet	CP dataset	Accuracy 93.97%
11	Ys et al. (2025)	Acoustic and textual features	ALS	Accuracy 81.8%
12	Javanmardi et al. (2024)	Pre-trained models	UA-Speech	Accuracy 68.25%
13	Yang et al. (2025)	Multi-head attention, multi-task	TORGO	Accuracy 62.5%
14	Gupta et al. (2022)	DNN architectures	UA-Speech	Accuracy 93.49%
15	Kim et al. (2021)	AFM signal model	Custom dataset	Accuracy 79.44%
16	Bhat & Strik (2025)	Two-stage data augmentation	Dysarthric ASR	Accuracy 75.80%
17	Varma et al. (2024)	Enhanced detection in CP and ALS	Biomedical signals	Accuracy 93.97%

End of table 1

No	Author(s) (Year)	Method	Dataset	Main Result
18	Suresh et al. (2024)	Enhancing dysarthria severity	SN Applied Sciences	Accuracy 70%
19	Joshy & Rajan (2021)	Automated dysarthria severity	IEEE Access	Accuracy 93.97%
20	Aljarallah et al. (2022)	Image classification-driven	Speech Communication	Accuracy 81.8%
21	Shabber & Sumesh (2024).	AFM signal model for dysarthric speech	PMC	Accuracy 77.76%
22	Joshy & Rajan (2023)	Dysarthria severity classification	Speech Communication	Accuracy 68.25%
23	Al-Ali et al. (2023)	Classification of Dysarthria	arXiv	Accuracy 86.9%
24	Joshy, A. A., & Rajan, R. (2022).	Speaker-independent intelligibility	IEEE Access	Accuracy 79.44%
25	Stumpf et al. (2024)	Speaker-Independent Dysarthria	arXiv	Accuracy 75.80%
Note – compiled by the author				

This table shows the main contribution of the literature, mostly focused on deep learning and using UA-Speech, TORGO datasets. The shortcomings of the studies are the small number of data, noisy audio, and the low generalization ability of the models.

CNN ARCHITECTURE

In addition to classical machine learning models, a Convolutional Neural Network (CNN) was implemented to automatically extract discriminative features from speech signals. The input to the network consisted of MFCC feature maps with a dimension of 40×128×140 \times 128 \times 140×128×1. The proposed architecture included three convolutional layers with 32, 64, and 128 filters, respectively, each using a 3×33 \times 33×3 kernel and ReLU activation function. Each convolutional layer was followed by a 2×22 \times 22×2 max-pooling layer to reduce spatial dimensionality.

After the convolutional blocks, the feature maps were flattened and connected to a fully connected dense layer with 128 neurons. A dropout layer with a rate of 0.5 was applied to reduce overfitting. The final output layer used a sigmoid activation function for binary classification. The model was trained using the Adam optimizer with binary cross-entropy loss.

MATERIALS AND METHODS

In this research work, the stages of data collection, pre-processing, feature extraction, model training and evaluation of results were carried out sequentially. First, the widely used UA-Speech and TORGO datasets were taken as a database. While the UA-Speech corpus contains speech recordings of different levels of speakers with and without dysarthria, the TORGO dataset

contains articulatory movement signals in addition to sound samples. The simultaneous use of both databases increases the stability of the models and ensures the reliability of the overall conclusion. The architecture used in the study is presented in Figure 1.

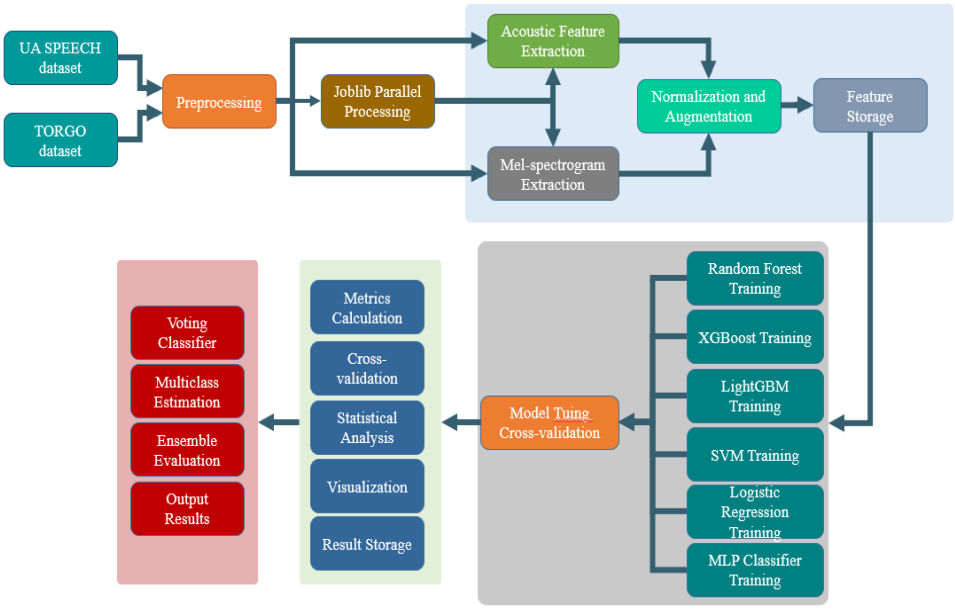

Figure 1. Architecture used in the study

Note – compiled by the author

All collected audio data underwent a series of preprocessing steps. At this stage, spectral gating was used to reduce random noise from audio signals, and all recordings were downsampled to 16 kHz. Amplitude values were normalized to the interval [−1,1], and poor-quality or too short (less than 0.5 seconds) recordings were removed. The Joblib library, which is based on parallel computing, was used to accelerate the processing of large amounts of data.

The main information characteristics of the study were extracted from the preprocessed signals. During feature extraction, several parameters were calculated that reflect the sound quality. These include the Melo-Frequency Cepstral Coefficients (MFCC), chroma features, spectral centroid, and the Zero-Crossing Rate of the signal. The MFCC coefficients were determined by the following formula:

$$c_n = \sum_{k=1}^{K} \log(S_k) * \cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{K}\right], \quad n = 1, \dots, N \tag{1}$$

Where S_k is the spectral coefficients calculated on the Mel scale, and K is the number of frequencies. In addition, the spectral centroid describes the average frequency value and is defined by the following expression:

$$C = \frac{\sum_f f * |X(f)|}{\sum_f |X(f)|} \tag{2}$$

Where $X(f)$ – is the amplitude values of the spectrum, f – is the frequency. All obtained features were normalized using the StandardScaler method and introduced into subsequent models.

Ten different machine learning models were considered in the study. Among them, along with traditional methods such as logistic regression, decision trees, SVM, KNN, ensemble methods such as Random Forest, Gradient Boosting, AdaBoost, XGBoost, LightGBM were also used. In addition, a neural network based on a multilayer perceptron (MLP Classifier) was

introduced, which allowed modeling nonlinear dependencies. Hyperparameters for each model were selected using the GridSearchCV method, and the dataset was stratified into 80% training and 20% test parts.

The results of the models were evaluated according to several metrics. Accuracy was calculated using the following formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

and precision and recall, respectively

$$Precision = \frac{TP}{TP+FP}, \quad Recall = \frac{TP}{TP+FN} \quad (4)$$

The F1-dimension was considered a harmonious balance between accuracy and completeness:

$$F1 = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (5)$$

These metrics allowed us to objectively assess the ability of models to work with unbalanced data. In addition, the error matrix, ROC-AUC curves, and precision-recall diagrams were created and the results were visualized.

At the final stage, the models that showed the highest results were combined based on the Voting Classifier. Using the soft voting method, the probability values of each model were combined and a final decision was made. This approach reduced the probability of error and increased stability compared to individual models. As a result, the ensemble approach showed higher F1-score and ROC-AUC indicators than individual models.

RESULTS

This study compared the performance of several machine learning models on the UA-Speech and TORGO speech datasets and evaluated their effectiveness in automatically detecting dysarthria. The main metrics used were accuracy, precision, recall, and F1-score. The models were evaluated after imbalance correction using ADASYN and hyperparameter optimization using GridSearchCV (Table 2).

Table 2. Model performance (%) on UA-Speech and TORGO datasets

Dataset / Model	UA-Speech				TORGO			
	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
RandomForest	92.40	96.73	78.11	86.43	89.25	88.67	90.00	89.33
XGBoost	94.85	95.85	87.17	91.30	93.75	93.53	94.00	93.77
LightGBM	94.74	96.61	86.04	91.02	93.50	93.07	94.00	93.53
SVM	92.63	92.08	83.40	87.52	92.50	91.26	94.00	92.61
Logistic Regression	79.42	65.61	70.57	68.00	96.25	95.57	97.00	96.28
MLPClassifier	91.46	91.38	80.00	85.31	95.50	93.75	97.50	95.59
Note – compiled by the author								

Based on the overall results, the boosting-based XGBoost and LightGBM models were identified as the most stable and high-performing models for dysarthria detection. These models proved to generalize well to complex acoustic features and effectively distinguish between different speech patterns.

The confusion matrix, expressed in percentages, of the Logistic Regression model used to detect dysarthria on the TORGO dataset is presented. The matrix shows how well and how badly the model classified into two classes - Control (healthy speech) and Dysarthric (dysarthric speech) (Figure 2).

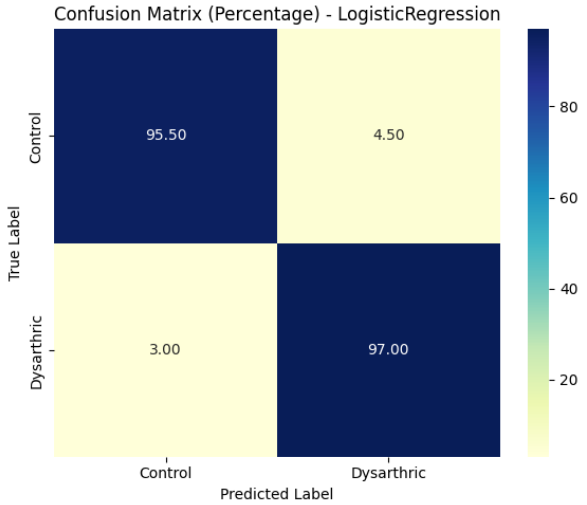

Figure 2. Confusion Matrix of the Logistic Regression Model – Percentages

Note – Compiled by the author

The matrix results show that the model performed very well:

95.50% of Control samples were correctly identified, while only 4.50% were misclassified as dysarthric. In addition, 97.00% of Dysarthric samples were correctly identified, while 3.00% were misclassified as Control.

These results show that the Logistic Regression model achieves very high recall and precision when working with TORGO data. High values on the diagonal of the matrix indicate that the model is reliable in recognizing dysarthric speech.

The confusion matrix of the XGBoost model used on the UA-Speech or TORGO dataset is shown. The matrix shows the numerical values of how well and how badly the model classified the two classes – 0 (Control) and 1 (Dysarthric) – (Figure 3).

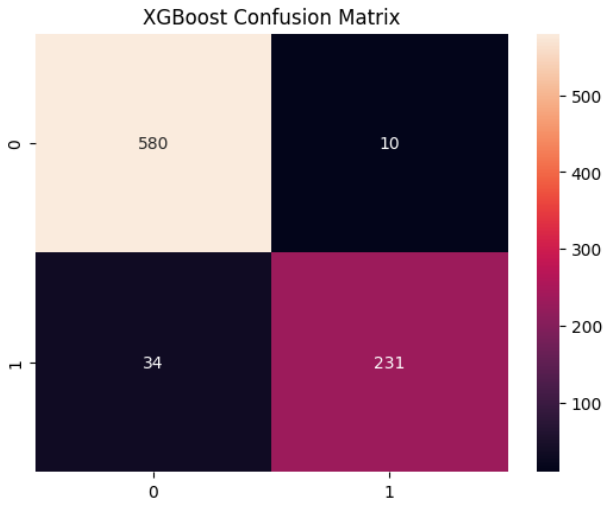

Figure 3. Confusion Matrix of the XGBoost Model

Note – Compiled by the author

The results of the matrix demonstrate the high performance of the XGBoost model:

- 580 samples belonging to the Control (0) class were correctly identified, and only 10 samples were misclassified as dysarthric.

- 231 samples belonging to the Dysarthric (1) class were correctly identified, and 34 samples were misclassified as Control.

These results indicate that the overall accuracy of the XGBoost model is high, especially for the Control class, and that it performs very consistently. The model also discriminates the Dysarthric class well, but due to the inherent complexity of the data, some samples are misclassified.

CONCLUSION

This article comprehensively analyzes the effectiveness of machine learning and deep learning models aimed at automatic dysarthria detection. RandomForest, XGBoost, LightGBM, SVM, Logistic Regression, MLPClassifier, and CNN models were compared on international speech datasets such as UA-Speech and TORGO, and their performance was evaluated in terms of accuracy, precision, recall, and F1-measure.

The experiments conducted showed that the boosting-based XGBoost and LightGBM models achieved the highest results in dysarthria detection. These models were able to effectively analyze complex acoustic features and distinguish pathological changes in speech signals with high accuracy. The SVM and MLPClassifier models also showed stable and competitive results, especially on the TORGO data. Logistic Regression, despite its simple structure, in some cases (especially for the TORGO data) gave very high results, demonstrating the impact of the structure and features of the data on the model performance.

The effectiveness of the ADASYN and class weight methods in solving the problem of unbalanced data was proven. These approaches increased the sensitivity of the models and improved the ability to detect dysarthria patterns. The resulting confusion matrices allowed for a deeper understanding of the behavior of the models, indicating areas of error. Overall, the results of the study demonstrated the high potential of machine learning methods in the automatic detection of dysarthria. These technologies contribute to accelerating the clinical diagnostic process, reducing subjectivity, and expanding the possibilities of early detection. The results of the study can serve as an important scientific basis for the development of medical decision support systems, telemedicine platforms, and automated speech therapy tools in the future.

CONFLICTS OF INTEREST: The authors declare no conflicts of interest.

FUNDING: No funding was received for the development of this article.

INSTITUTIONAL REVIEW BOARD STATEMENT: This study was conducted using publicly available UA-Speech and TORGO speech datasets and focused on the comparative evaluation of machine learning and deep learning models for dysarthria detection. The authors did not conduct any direct experiments involving human participants or animals. Therefore, approval from an Institutional Review Board or an equivalent ethics committee was not required.

INFORMED CONSENT STATEMENT: No human participants were directly recruited or involved by the authors in this study. Since the research was based on previously available speech datasets, additional written or oral informed consent was not required within the scope of this work.

DATA AVAILABILITY STATEMENT: The data supporting the findings of this study are based on the publicly available UA-Speech and TORGO speech datasets. The processed data, model training results, and analytical materials generated during the study are available from the corresponding author upon reasonable request.

ACKNOWLEDGEMENTS: The authors express their sincere gratitude to the «AI RESEARCHER LAB» laboratory of Khoja Akhmet Yassawi International Kazakh-Turkish University for providing experimental equipment, technical support, and valuable assistance during the research work.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES: No AI technologies were used at any stage in the development of this article.

REFERENCES

Jayaraman, D. K., & Das, J. (2023). Dysarthria. In StatPearls [Internet]. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK592453

Prabhala, J. C., Ragoju, R., Kupilli, V., & Chesneau, C. (2025). Enhanced early detection of dysarthric speech disabilities using stacking ensemble deep learning model. Machine Learning with Applications, 21, 100721. https://doi.org/10.1016/j.mlwa.2025.100721

Smyrni, V., Giannopapas, V., Kitsos, D. K., Stavrogianni, K., Chasiotis, A. K., Papagiannopoulou, G., ... & Giannopoulos, S. (2025). Prevalence of dysarthria in the multiple sclerosis population: A systematic review and meta-analysis. Multiple sclerosis and related disorders, 98, 106458. https://doi.org/10.1016/j.msard.2025.106458

Cerebral Palsy Alliance. (n.d.). Dysarthria and cerebral palsy fact sheet. https://cpresource.org/topic/communication/dysarthria-and-cerebral-palsy-fact-sheet

National Institute on Deafness and Other Communication Disorders. (n.d.). Quick statistics about voice, speech, and language. https://www.nidcd.nih.gov/health/statistics/quick-statistics-voice-speech-language

Lin, L.-X., Yao, S.-Y., Chen, Q., Du, M.-Q., Kang, X.-H., Jiang, D.-Y., & Peng, Y. (2025). Stroke-associated dysarthria. Frontiers in Neurology, 16, Article 1629640. https://doi.org/10.3389/fneur.2025.1629640

Hassan, E., Saber, A., Abd El-Hafeez, T., Medhat, T., & Shams, M. Y. (2025). Enhanced dysarthria detection in cerebral palsy and ALS patients using WaveNet and CNN-BiLSTM models: A comparative study with model interpretability. Biomedical Signal Processing and Control, 110, 108128. https://doi.org/10.1016/j.bspc.2025.108128

Shih, D. H., Liao, C. H., Wu, T. W., Xu, X. Y., & Shih, M. H. (2022). Dysarthria speech detection using convolutional neural networks with gated recurrent unit. Healthcare, 10(10), 1956. https://doi.org/10.3390/healthcare10101956

Bhat, C., & Strik, H. (2025). Speech technology for automatic recognition and assessment of dysarthric speech: An overview. Journal of Speech, Language, and Hearing Research, 68(2), 547-577. https://doi.org/10.1044/2024_JSLHR-23-00740

Varma, V. J., Jana, A., Samal, A. K., S, A., Naidu, R. C., Al-Maadeed, S., ... & Khoodeeram, R. (2025). Enhancing dysarthria severity classification: efficient audio based deep learning models. Discover Applied Sciences, 7(8), 886. https://doi.org/10.1007/s42452-025-07260-2

Stuart, C. M. (2025, October). Comparative Analysis of Deep Learning Models for Dysarthria Severity Detection using Balanced Speech Data. In 2025 International Conference on Power, Instrumentation, Control, and Computing (PICC) (pp. 1-6). IEEE. https://doi.org/10.1109/PICC67314.2025.11291205

Suresh, M., Rajan, R., & Thomas, J. (2025). Speaker independent dysarthria severity classification using synthesis-based augmentation. International Journal of Speech Technology, 28(2), 565-580. https://doi.org/10.1007/s10772-025-10191-3

Wihlborg, L., Goodall, J., Wheatley, D., Webber, J. J., Tam, J., Weaver, C., ... & Valentini-Botinhao, C. (2025). Evaluating pretrained speech embedding systems for dysarthria detection across heterogenous datasets. arXiv preprint arXiv:2509.19946. https://doi.org/10.48550/arXiv.2509.19946

Vogel, A. P., Graf, L., Gauß, M., Chan, C. S., Hepworth, G., & Synofzik, M. (2025). Development and validation of the dysarthria impact scale. medRxiv, 2025-07. https://doi.org/10.1101/2025.07.02.25330604

Alharbi, G. F., Alamri, N. K., & Sabbeh, S. F. (2025). Automatic classification of speech Dysarthric intelligibility levels using textual feature. IEEE Access, 13, 39982-39992. https://doi.org/10.1109/ACCESS.2025.3547041.

Al-Ali, A., Al-Maadeed, S., Saleh, M., Naidu, R. C., Alex, Z. C., Ramachandran, P., & Khoodeeram, R. (2023). Classification of dysarthria based on the levels of severity. A systematic review. arXiv preprint arXiv:2310.07264. https://doi.org/10.48550/arXiv.2310.07264

Bhat, C., & Strik, H. (2025). Two-stage data augmentation for improved ASR performance for dysarthric speech. Computers in Biology and Medicine, 189, 109954. https://doi.org/10.1016/j.compbiomed.2025.109954

Joshy, A. A., & Rajan, R. (2021, January). Automated dysarthria severity classification using deep learning frameworks. In 2020 28th European signal processing conference (EUSIPCO) (pp. 116-120). IEEE. https://doi.org/10.23919/Eusipco47968.2020.9287741.

Aljarallah, N. A., Dutta, A. K., & Sait, A. R. W. (2025). Image classification-driven speech disorder detection using deep learning technique. SLAS technology, 32, 100261. https://doi.org/10.1016/j.slast.2025.100261

Shabber, S. M., & Sumesh, E. P. (2024). AFM signal model for dysarthric speech classification using speech biomarkers. Frontiers in Human Neuroscience, 18, 1346297. https://doi.org/10.3389/fnhum.2024.1346297

Joshy, A. A., & Rajan, R. (2023). Dysarthria severity classification using multi-head attention and multi-task learning. Speech Communication, 147, 1-11. https://doi.org/10.1016/j.specom.2022.12.004

Javanmardi, F., Kadiri, S. R., & Alku, P. (2024). Pre-trained models for detection and severity level classification of dysarthria from speech. Speech Communication, 158, 103047. https://doi.org/10.1016/j.specom.2024.103047

Yang, Y., Liang, R., Ni, Y., Xie, Y., Zou, C., & Schuller, B. W. (2025). A Non-Intrusive Speech Quality Evaluation Framework for Hearing Aids Based on Speech Label Assistance and Multi-Task Learning Strategy. IEEE Transactions on Audio, Speech and Language Processing.
https://doi.org/10.1109/TASLPRO.2025.3594293

Gupta, M., Bhargava, L., & Indu, S. (2022). Deep neural network learning for power limited heterogeneous system with workload classification. computing, 104(1), 95-122. https://doi.org/10.1007/s00607-021-01018-5

Kim, Y. M., Lee, J., Jeon, D. J., Oh, S. E., & Yeo, J. S. (2021). Advanced atomic force microscopy-based techniques for nanoscale characterization of switching devices for emerging neuromorphic applications. Applied Microscopy, 51(1), 7. https://doi.org/10.1186/s42649-021-00056-9

Merler, M., Agurto, C., Peller, J., Roitberg, E., Taitz, A., Trevisan, M. A., ... & Norel, R. (2025). Clinical assessment and interpretation of dysarthria in ALS using attention based deep learning AI models. npj Digital Medicine, 8(1), 260. https://doi.org/10.1038/s41746-025-01654-7

Shahamiri, S. R., Lal, V., & Shah, D. (2023). Dysarthric speech transformer: A sequence-to-sequence dysarthric speech recognition system. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, 3407-3416. https://doi.org/10.1109/TNSRE.2023.3307020.

Anuprabha, M., Gurugubelli, K., Kesavaraj, V., & Vuppala, A. K. (2025, April). A multi-modal approach to dysarthria detection and severity assessment using speech and text information. In ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and

Signal Processing (ICASSP) (pp. 1-5). IEEE. https://doi.org/10.1109/ICASSP49660.2025.10889515.

Stumpf, L., Kadirvelu, B., Waibel, S., & Faisal, A. A. (2024). Speaker-independent dysarthria severity classification using self-supervised transformers and multi-task learning. arXiv preprint arXiv:2403.00854. https://doi.org/10.48550/arXiv.2403.00854

Joshy, A. A., & Rajan, R. (2022). Automated dysarthria severity classification: A study on acoustic features and deep learning techniques. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 30, 1147-1157. https://doi.org/10.1109/TNSRE.2022.3169814

YS, U. V., Bhattacharjee, T., Udupa, S., Chowdam Venkata Thirumala, K., Keerthipriya, M., Chikktimmegowda, D., ... & Ghosh, P. K. (2025). Comparison of Acoustic and Textual Features for Dysarthria Severity Classification in Amyotrophic Lateral Sclerosis. In Proc. Interspeech 2025 (pp. 803-807). https://www.isca-archive.org/interspeech_2025/ys25_interspeech.pdf

Авторлар туралы мәліметтер
Информация об авторах
Information about authors

Абен Арыпжан Бактиярович – магистр, Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті, Түркістан қ., Қазақстан,

Абен Арыпжан Бактиярович – магистр, Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, г.Туркестан, Казахстан

Aben Arypzhan Baktiarovich - master, International Kazakh-Turkish University named after Khoja Ahmed Yasawi, Turkestan, Kazakhstan
e-mail: arypzhan.aben@ayu.edu.kz
ORCID: https://orcid.org/0000-0001-8534-3288

Мамырбаев Оркен Жумажанович – PhD, профессор, Ақпараттық және есептеу технологиялары институты, Алматы қ., Қазақстан

Мамырбаев Оркен Жумажанович – PhD, профессор, Института информационных и вычислительных технологий, г. Алматы, Казахстан

Mamyrbayev Orken Zhumazhanovich – PhD, Professor, Institute of Information and Computational Technologies, Almaty, Kazakhstan
e-mail: morkenj@mail.ru
ORCID: https://orcid.org/0000-0001-8318-3794

Казбекова Гулнур Нагиметовна – техника ғылымдарының кандидаты, қауымдастырылған профессор, Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті, Түркістан қ., Қазақстан
Казбекова Гулнур Нагиметовна – кандидат технических наук, доцент, Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, г. Туркестан, Казахстан
Kazbekova Gulnur Nagimetovna – Candidate of Technical Sciences, Associate Professor, International Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkestan, Kazakhstan
e-mail: gulnur.kazbekova@ayu.edu.kz
ORCID: https://orcid.org/0000-0002-2756-7926

Жунисов Нурсейт Мухидинович – PhD, аға оқытушы, Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті, Түркістан қ., Қазақстан
Жунисов Нурсейт Мухидинович – PhD, старший преподаватель, Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, г. Туркестан, Казахстан,
Zhunisov Nurseit Mukhidinovich – PhD, Khoja Ahmed Yasawi International Kazakh-Turkish University, Turkestan
e-mail: nurseit.zhunissov@ayu.edu.kz,
ORCID: https://orcid.org/0000-0001-6531-9408

Баймаханова Айгерим Саттаровна – PhD, аға оқытушы, *Қожа* Ахмет Ясауи атындағы Халықаралық қазақ-түрік *университеті*, Түркістан қ., Қазақстан
Баймаханова Айгерим Саттаровна – PhD, старший преподаватель, Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, г. Туркестан, Казахстан
Baimakhanova Aigerim Sattarovna – PhD, Senior Lecturer, Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan, Kazakhstan
e-mail: aygerim.baymakhanova@ayu.edu.kz
ORCID: https://orcid.org/0000-0002-5364-0146