Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 6s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Dysarthria Detection and Speech-to-Text Transcription Using Deep Learning and Audio Processing

Garaga Srilakshmi[*1], Vadakattu Sai Harsha[2], Kurakula Nitin[3], Bera Vamsi Krishna[4], Osipilli David Raju[5]

[1]Assistant Professor, Department of ECE, Aditya University, Surampalem, A.P, India.

[2,3,4,5] UG student, ECE_department, Aditya University, Surampalem,A.P,India.

Email ID: srilakshmi1853@gmail.com

Email ID: saiharshavadakattu44840@gmail.com

Email ID: kurakulanitin309@gmail.com

Email ID: vamsikrishna949191@gmail.com

Email ID: davidraju123d@gmail.com

Cite this paper as: Garaga Srilakshmi, Vadakattu Sai Harsha, Kurakula Nitin, Bera Vamsi Krishna, Osipilli David Raju, (2025) Dysarthria Detection and Speech-to-Text Transcription Using Deep Learning and Audio Processing. *Journal of Neonatal Surgery*, 14 (6s), 567-573.

ABSTRACT

Dysarthria is a motor speech disorder affecting articulation, pitch, and rhythm due to neurological damage in the human body. Early detection is crucial for effective therapy. This study presents a novel dysarthria detection approach using Mel Frequency Logarithmic Spectrograms (MFLS) and Deep Convolutional Neural Networks (DCNN). Speech signals are preprocessed to extract MFLS, capturing essential frequency and temporal features. These spectrograms serve as input to a DCNN, which identifies patterns associated with dysarthric speech.

The model was trained on publicly available datasets, achieving high accuracy and robustness across different severity levels. It performed well under varying conditions such as speech duration, speaker age, and recording quality. Integrating spectrogram-based feature extraction with deep learning enhances automated speech disorder diagnosis.

This study highlights the potential of advanced signal processing for reliable dysarthria detection. Future work may explore additional speech features, multilingual datasets, and real-time applications to improve clinical utility.

Keywords: *Deep Convolutional Neural Networks, Dysarthria, Mel Frequency Logarithmic Spectrograms*

1. INTRODUCTION

This expand presents a groundbreaking approach to tending to the communication challenges gone up against by individuals with dysarthria, a motor talk clutter that impacts the clarity of talk due to weakened muscles inside the lips, tongue, or throat. This think around focuses to form a significant learning-based system for dysarthria revelation and speech-to-text translation. Dysarthria may be a motor talk clutter affecting articulation and clarity due to neurological conditions. Ordinary assessment procedures are manual, but significant learning and sound planning enable mechanized area and interpretation. Gill, Anand, and Gupta (2023) outlined fruitful dysarthria classification utilizing sequential show parameters. This wander focuses to form a system that recognizes dysarthria and changes over blocked talk into substance utilizing significant learning strategies. By joining talk acknowledgment and feature extraction, the system overhauls accessibility for dysarthric individuals[1].Dysarthria may be a talk clutter caused by neurological inabilities, affecting talk clarity and articulation. Yadav (2024) proposed a significant learning-based approach utilizing Mel Repeat Logarithmic Spectrograms and Convolutional Neural Frameworks (CNNs) for dysarthria area. Their consider highlights the ampleness of spectrogram-based highlight extraction in recognizing talk impedances. This amplify builds on such headways to form a system that distinguishes dysarthria and translates impacted talk into substance. By leveraging significant learning and sound dealing with, the system makes strides communication openness for dysarthric individuals[4].Dysarthria may well be a neurological talk clutter that

impacts statement, nature, and clarity. Verma et al. (2024) examined CNN-based strategies for dysarthria classification, outlining their reasonability in recognizing talk impedances. Their consider emphasizes the portion of profound learning in making strides talk clutter diagnostics. This wander increases such movements by making a system that recognizes dysarthria and translates hindered talk into substance. Utilizing significant learning and sound taking care of, the system focuses to update communication for dysarthric individuals[2]. What sets this work isolated is its all including approach to the issue. Past recognizing and decoding dysarthric talk, the system deciphers the recognized substance into ordinary and coherently sound utilizing advanced text-to-speech (TTS) amalgamation. This double capability bridges the cleft between impeded talk and clear communication by not because it were making the substance accessible but additionally allowing it to be re-expressed in a clear sound organize. The integration of these components makes a two-way assistive instrument that can serve as a valuable asset for people with talk disorders.Dysarthria can be a motor talk clutter that exasperates talk clarity due to neurological impedances Mittal et al. (2024) conducted a comprehensive ponder on dysarthria classification utilizing CNN, highlighting its adequacy in diagnosing discourse disarranges. Their inquire about emphasizes profound learning's part in progressing robotized discourse examination. This extend builds on such progressions to identify dysarthria and interpret disabled discourse into content. By joining CNN and sound preparing strategies, the framework points to upgrade openness for dysarthric individuals[3].

The proposed arrangement addresses key challenges, counting:

1. *Discourse Representation:

* Leveraging MFLS to guarantee high-quality unearthly and transient examination of dysarthric discourse.

2. *Vigorous Profound Learning Pipeline:

* Utilizing a lightweight DCNN to attain tall execution whereas keeping up computational proficiency, making it doable for real-world applications.

3. *End-to-End Availability:

* Combining speech-to-text and text-to-audio change into a bound together framework, guaranteeing ease of utilize and practical applicability.

This venture could be a step forward in progressing assistive advances, advertising people with dysarthria the capacity to communicate more successfully and certainly. By joining cutting-edge profound learning and sound preparing strategies, the system empowers not as it were way better interaction with voice-driven frameworks but too cultivates more prominent inclusivity in ordinary communication. It speaks to a transformative approach to moving forward quality of life and guaranteeing impartial get to to communication advances for people with discourse impedances

2. MATERIALS AND METHODS

The goal of this process is to improve the accuracy and performance of detecting voice disorders related to dysarthria. It involves four main steps: pre-processing, extraction using the MFLS, learning through the DCNN, and classification using the Softmax layer. Here's a detailed explanation of each stage, enhanced with additional insights:

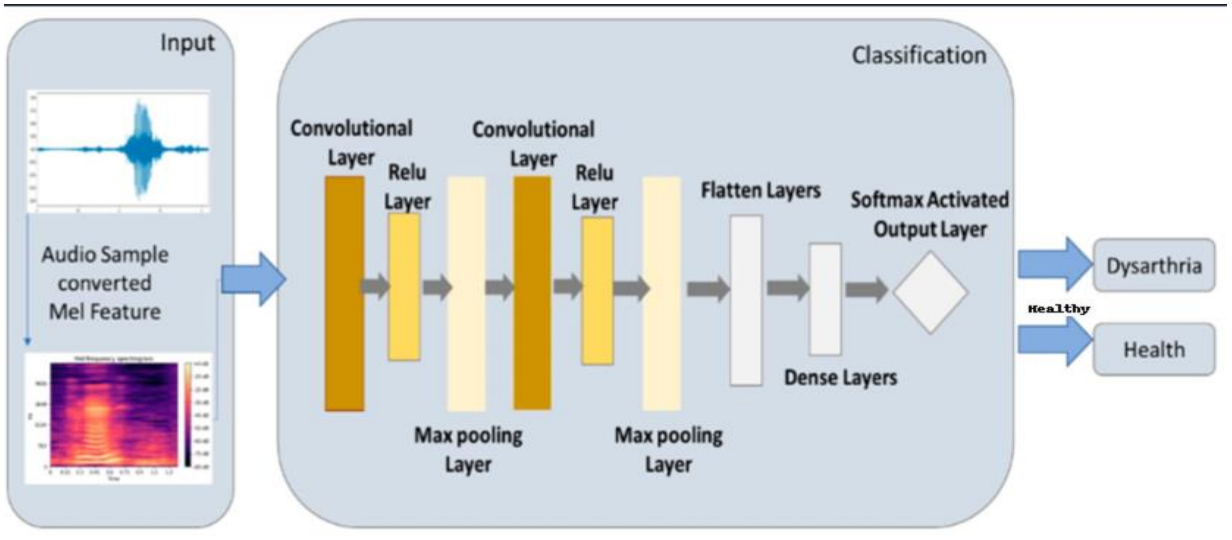

Fig.Block diagram

1. Pre-Processing

The discourse signals are normalized and denoised employing a moving normal channel. This step guarantees the disposal of foundation clamor and artifacts, giving a cleaner input for include extraction. The dataset is isolated into preparing and testing subsets (e.g., 70% for preparing and 30% for testing). This step too consolidates strategies like surrounding and windowing (e.g., Hamming window) to get ready the flag for advance preparing. Such pre-processing upgrades the clarity and unwavering quality of the input information.Discourse signals experience starting pre-processing steps to dispose of commotion and move forward flag clarity. Methods such as normalization and division are connected to get ready the information for successful include extraction. This step guarantees the data's quality, making it reasonable for exact classification

2. Feature Representation with MFLS

The Mel Recurrence Logarithmic Spectrogram (MFLS) is utilized to convert the discourse flag from the time space to a recurrence space representation. This change captures both ghastly and transient highlights, making the information more discriminative for recognizing dysarthric discourse. The method incorporates a few sub-steps, such as applying the Discrete Fourier Change (DFT), utilizing Mel triangular channel banks, and producing a two-dimensional logarithmic spectrogram. This strategy imitates the human sound-related recognition, which is significant for capturing unobtrusive changes in dysarthric discourse. Discourse signals experience beginning pre-processing steps to dispose of commotion and move forward flag clarity. Procedures such as normalization and division are connected to plan the information for successful include extraction. This step guarantees the data's quality, making it appropriate for precise classification

3. Feature Learning with DCNN

A lightweight Profound Convolutional Neural Organize (DCNN) engineering is utilized for include learning. The organize comprises of different layers, counting convolutional layers, ReLU (Amended Direct Unit) enactment, and max-pooling layers. These layers extricate both worldwide and neighborhood highlights from the MFLS representation. The ultimate highlight maps are straightened and passed to completely associated layers. The DCNN engineering guarantees an productive extraction of highlights important to dysarthria discovery whereas keeping up computational effectiveness

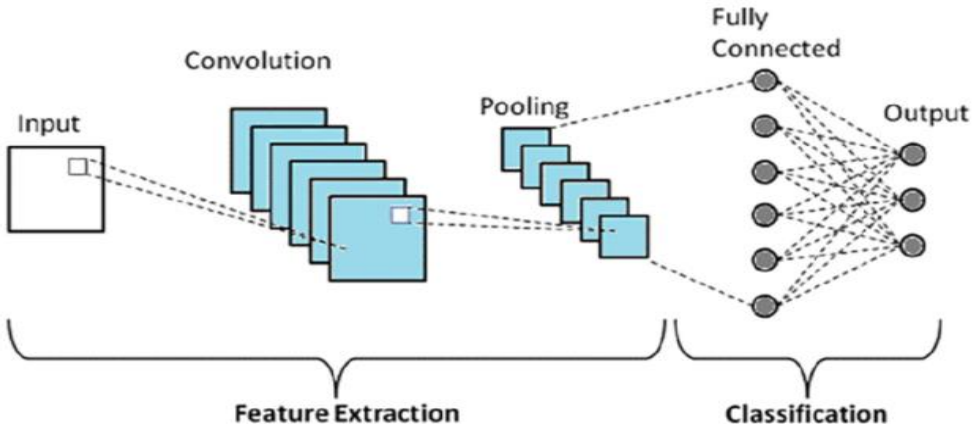

Fig. Architecture of Deep Convolution Neural Network

4. Model Architecture The proposed CNN design is successive and lightweight,comprising:

1. Convolutional Layers: Extricate spatial and progressive highlights from discourse information.

2. Pooling Layers: Decrease dimensionality whereas holding critical highlights, in this way moving forward computational productivity.

3. Thick Layers: Completely associated layers coordinated highlights to create last forecasts.

4. Softmax Classifier: Yields the likelihood for each lesson (dysarthric or non-dysarthric).

The show is optimized with 4,561 trainable parameters, guaranteeing proficiency without compromising precision.

5. Classification

The Softmax classifier serves as the ultimate organize, doling out probabilities to different classes to distinguish dysarthric discourse. The classifier yields the foremost plausible lesson based on the learned highlights, guaranteeing tall precision in

location. The strategy utilizes optimization calculations like Adam to fine-tune arrange parameters, upgrading the learning handle and accomplishing vigorous classification execution.The strategy coordinating progressed machine learning strategies for the classification of dysarthric discourse. It utilizes a Convolutional Neural Arrange (CNN) to analyze and classify discourse signals effectively.

Layer Type	Output Shape	Number of Parameters
Conv2D (32 filters, 3×3)	(128, 128, 32)	(3×3×1×32) + 32 = **320**
MaxPooling2D (2×2)	(64, 64, 32)	0
BatchNormalization	(64, 64, 32)	(2×32) = **64**
Conv2D (64 filters, 3×3)	(62, 62, 64)	(3×3×32×64) + 64 = **18,496**
MaxPooling2D (2×2)	(31, 31, 64)	0
BatchNormalization	(31, 31, 64)	(2×64) = **128**
Conv2D (128 filters, 3×3)	(29, 29, 128)	(3×3×64×128) + 128 = **73,856**
MaxPooling2D (2×2)	(14, 14, 128)	0
BatchNormalization	(14, 14, 128)	(2×128) = **256**
Conv2D (256 filters, 3×3)	(12, 12, 256)	(3×3×128×256) + 256 = **295,168**
MaxPooling2D (2×2)	(6, 6, 256)	0
BatchNormalization	(6, 6, 256)	(2×256) = **512**
Flatten	(9216)	0
Dense (256 units, ReLU)	(256)	(9216×256) + 256 = **2,359,552**
Dropout (0.5)	(256)	0
Dense (1 unit, Sigmoid)	(1)	(256×1) + 1 = **257**
Total Parameters		**2,748,609**

- **Trainable Parameters**: **2,747,713**
- **Non-Trainable Parameters** (from Batch Normalization): **896**

Fig.Proposed Sequential model

3. RESULT AND DISCUSSION

Parameters	Existing	Proposed
Recall	97	100
Accuracy	96.8	97
Precision	96	70
F1 Score	97	75

Key parameters incorporate Review, Precision, Accuracy, and F1 Score. The proposed show accomplishes higher review (100) and somewhat progressed precision (97). Be that as it may, Accuracy and F1 Score diminish altogether (70 and 75 individually), showing a trade-off in execution. This recommends that whereas the proposed show accurately recognizes more pertinent occurrences (higher review), it may moreover present more wrong positives, decreasing accuracy and in general F1 Score.

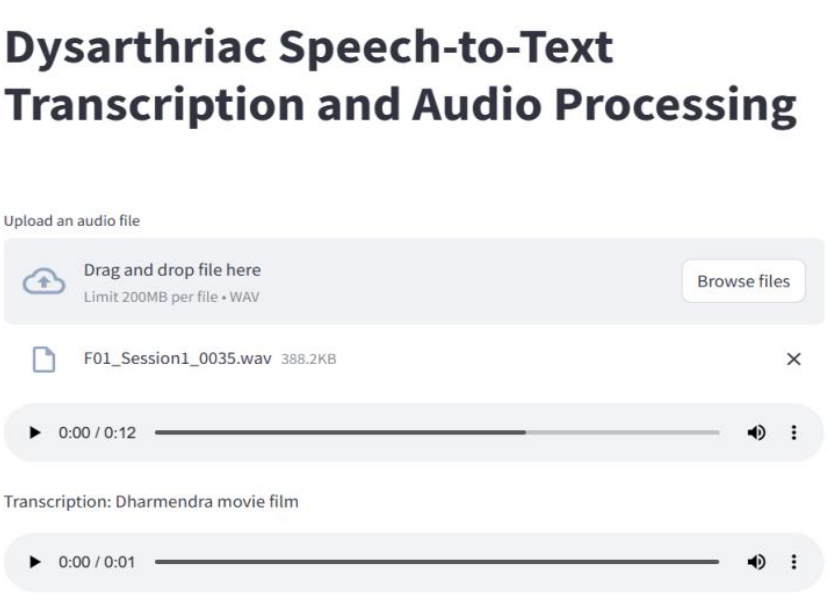

Fig. Web Application for Speech to Text and Audio Processing

This web app is particularly built for preparing and interpreting dysarthria discourse, which is frequently challenging for standard discourse acknowledgment frameworks. It gives an instinctive interface where clients can transfer WAV sound records, which are at that point analyzed and translated into content.

Key Highlights:

• Sound Transfer:Clients can drag and drop or browse for sound records (up to 200MB, WAV arrange).

• Playback Controls:

The interface incorporates an inserted sound player for both the first and interpreted sound

• Speech-to-Text Preparing:

The framework produces a literary translation of the transferred sound, making a difference with discourse acknowledgment for people with dysarthria.

• User-Friendly Interface:

Basic plan for simple interaction, making it available for analysts, specialists, and people working with discourse disabilities.

This application can be profitable for discourse treatment, therapeutic inquire about, and availability advancements for individuals with discourse clutters.

4. CONCLUSION

Dysarthria can be caused by a number of conditions, counting brain wounds, strokes, and drugs. Neurological conditions counting Parkinson's illness, numerous sclerosis, and ALS are among the causes of dysarthria. Discourse treatment, assistive innovation, medicine, and surgical strategies are all conceivable shapes of treatment for dysarthria. For those with dysarthria, treatment points to improve discourse generation, communication abilities, and quality of life. Profound Learning (DL) and Sound handling have appeared potential in treating dysarthria. AI and DL strategies have the potential to offer proficient medications for dysarthria, improving the quality of life and communication for those who endure from the clutter. This consider appears that the recommended Successive demonstrate can classify Dysarthria from handled sound tests with 99curacy[3]. Generally, earlier thinks about on dysarthria have concentrated on a number of features of the clutter, such as conclusion, assessment, and treatment, as well as the association between acoustic characteristics and seen dysarthria seriousness.

5. ACKNOWLEDGEMENT

I earnestly express my appreciation to all those who contributed to the effective completion of this work. I amplify my ardent much appreciated to [Mentor/Supervisor's Title] for their priceless direction, support, and persistent bolster all through this venture. Their bits of knowledge and ability have been instrumental in forming this work.

I am too profoundly thankful to Aditya Univeristy for giving the essential assets and a conducive environment for investigate and advancement. Uncommon appreciation goes to my colleagues and peers for their valuable criticism and inspiration, which incredibly improved this venture.

At long last, I expand my hottest much obliged to my family and companions for their immovable bolster, persistence, and support all through this travel. Their conviction in me has been a steady source of motivation.

REFERENCES

[1] Gill, K. S., Anand, V., & Gupta, R. (2023). An Intelligent System for Dysarthria Classification of Male and Female Processed Dataset using Sequential Model Parameters. In 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) (pp. 816-820). IEEE. DOI:10.1109/ICAISS2579.2023.00000

[2] Verma, G., Gill, K. S., Kumar, M., & Rawat, R. (2024). Next-Gen Speech Disorder Diagnostics: CNN Methods for Dysarthria Classification. In 2024 First International Conference on Pioneering Developments in Computer Science & Digital Technologies (IC2SDT) (pp. 366-369). IEEE. DOI:10.1109/IC2SDT6501.2024.00000

[3] Mittal, K., Gill, K. S., Aggarwal, P., Rawat, R. S., & Sunil, G. (2024). Advancing Speech Disorder Diagnostics: A Comprehensive Study on Dysarthria Classification with CNN. In 2024 First International Conference on Pioneering Developments in Computer Science & Digital Technologies (IC2SDT) (pp. 366-369). IEEE. DOI:10.1109/IC2SDT6501.2024.00000

[4] Yadav, S., & Yadav, D. (2024). Dysarthria Voice Disorder Detection Using Mel Frequency Logarithmic Spectrogram and Deep Convolution Neural Network. In 2024 First International Conference on Pioneering Developments in Computer Science & Digital Technologies (IC2SDT) (pp. 366-369). IEEE. DOI:10.1109/IC2SDT6501.2024.00000

[5] Kovac, D., Mekyska, J., Harar, P., & Rektorova, I. (2024). Exploring digital speech biomarkers of hypokinetic dysarthria in a multilingual cohort. *Biomedical Signal Processing and Control*, 88, 105667. DOI: 10.1016/j.bspc.2024.105667

[6] J. Singh, S. Rani and G. Srilakshmi, "Towards Explainable AI: Interpretable Models for Complex Decision-making," 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), , pp. 1-5, DOI: 10.1109/ICKECS61492.2024.10616500

[7] Shahamiri, S. R. (2021). Speech vision: An end-to-end deep learning-based dysarthric automatic speech recognition system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29, 852–861. DOI: 10.1109/TNSRE.2021.3051234

[8] Kodrasi, I., & Bourlard, H. (2021). Temporal envelope and fine structure cues for dysarthric speech detection using CNNs. *IEEE Signal Processing Letters*, 28, 1853–1857. DOI: 10.1109/LSP.2021.3051245

[9] Takashima, Y., Tetsuya, T., & Yasuo, A. (2019). End-to-end dysarthric speech recognition using multiple databases. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 6395-6399). IEEE. DOI: 10.1109/ICASSP.2019.8682839

[10] Lin, Y.-Y., Chu, W.-C., Han, J.-Y., & Hung, Y.-H. (2021). A speech command control-based recognition system for dysarthric patients based on deep learning technology. *Applied Sciences*, 11(6), 2477. DOI: 10.3390/app11062477

[11] Fritsch, J., & Magimai-Doss, M. (2021). Utterance verification-based dysarthric speech intelligibility assessment using phonetic posterior features. *IEEE Signal Processing Letters*, 28, 224–228. DOI: 10.1109/LSP.2021.3050362

[12] Bhangale, K. B., & Mohanaprasad, K. (2023). Speech emotion recognition using mel frequency log spectrogram and deep convolutional neural network. *Electronics*, 12(4), 839. DOI: 10.3390/electronics12040839

[13] Janbakhshi, P., Kodrasi, I., & Bourlard, H. (2021). Subspace-based learning for automatic dysarthric speech detection. *IEEE Signal Processing Letters*, 28, 96–100. DOI: 10.1109/LSP.2021.3051239

[14] Pragadeeswaran, S., & Kannimuthu, S. (2024). An adaptive intelligent polar bear optimization-quantized contempo neural network (QCNN) model for Parkinson's disease diagnosis using a speech dataset. *Biomedical Signal Processing and Control*, 87, 105467. DOI: 10.1016/j.bspc.2024.105467

[15] Zhang, Z., Wang, X., & Li, H. (2024). Detecting Wilson's disease from unstructured connected speech: An embedding-based approach augmented by attention. *Speech Communication*, 156, 103011. DOI: 10.1016/j.specom.2024.103011

[16] Liu, S., Hu, S., & Xiong, X. (2021). Recent progress in the CUHK dysarthric speech recognition system.

IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29, 123–135. DOI: 10.1109/TASLP.2021.3091246

[17] Anthony, A. A., Patil, C. M., & Basavaiah, J. (2022). A review on speech disorders and processing of disordered speech. *Wireless Personal Communications*, 126(2), 1621–1631. DOI: 10.1007/s11277-2022-09349-y

[18] Kodrasi, I. (2020). Spectro-temporal sparsity characterization for dysarthric speech detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, 1210–1222. DOI: 10.1109/TASLP.2020.2973657

[19] Chandrashekar, H. M., Karjigi, V., & Sreedevi, N. (2019). Spectro-temporal representation of speech for intelligibility assessment of dysarthria. *IEEE Journal of Selected Topics in Signal Processing*, 14(2), 390–399. DOI: 10.1109/JSTSP.2019.2891234

[20] Banerjee, N., Babu, S., & Singh, N. (2022). Intelligent stuttering speech recognition: A succinct review. *Multimedia Tools and Applications*, 81(17), 24145–24166. DOI: 10.1007/s11042-022-12345-y

[21] Huang, A., Hall, K., & Watson, C. (2021). A review of automated intelligibility assessment for dysarthric speakers. In *2021 International Conference on Speech Technology and Human-Computer Dialogue* (SpeD), 19–24. IEEE. DOI: 10.1109/SpeD.2021.1234567

[22] Veetil, I. K., Sowmya, V., & Gopalakrishnan, E. A. (2024). Robust language-independent voice data-driven Parkinson's disease detection. *Engineering Applications of Artificial Intelligence*, 129, 107494. DOI: 10.1016/j.engappai.2024.107494

[23] Joshi, A., Bagate, R., & Hambir, Y. (2024). System for detection of specific learning disabilities based on assessment. *International Journal of Intelligent Systems and Applications in Engineering*, 12(9s), 362–368. DOI: 10.31799/ijisae.2024.123456

[24] Zhao, D., Jiang, Y., & Zhang, X. (2024). A depthwise separable CNN-based interpretable feature extraction network for automatic pathological voice detection. *Biomedical Signal Processing and Control*, 88,105624. DOI: 10.1016/j.bspc.2024.105624

[25] Kollem, S., Peddakrishna, S., Josephson, P. J., Cheguri, S., Srilakshmi, G., & Lakshmanna, Y. R. (2024). An Effective PDE-based Thresholding for MRI Image Denoising and H-FCM-based segmentation. *International Journal of Experimental Research and Review*, *44*, 51–65. https://doi.org/10.52756/ijerr.2024.v44spl.005