www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

VOICE DISORDER CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK BASED ON DEEP TRANSFER LEARNING

[1]V.VENKATA RAMANJANEYULU, [2]B. VAISHNAVI, [3]B. ARAVIND, [4]CH. MANIDEEP

[1] PROFESSOR, CSE, TEEGALA KRISHNA REDDY ENGINEERING COLLEGE

[234]UG. SCHOLAR, CSE, TEEGALA KRISHNA REDDY ENGINEERING COLLEGE

ABSTRACT

Voice disorders are prevalent medical conditions affecting individuals' ability to communicate effectively. Accurate and timely diagnosis of these disorders is crucial for proper treatment and management. In recent years, deep learning techniques, particularly convolutional neural networks (CNNs), have shown promising results in various classification tasks, including medical diagnostics. Transfer learning, a technique where a model trained on one task is adapted to another, has been particularly effective in scenarios where labeled data is limited. In this study, we propose a novel approach for voice disorder classification utilizing CNNs and deep transfer learning. Our methodology involves pretraining a CNN model on a large dataset of general audio samples to learn general acoustic features. We then fine-tune the pretrained model on a smaller dataset specific to voice disorder classification. This transfer learning process enables the model to leverage knowledge gained from the general dataset, improving its ability to classify voice disorders despite limited labeled data in the target domain.

I.INTRODUCTION

1.1 PROJECT INTRODUCTION

Voice disorders, encompassing a range of conditions that affect the quality, pitch, or volume of an individual's voice, pose significant challenges to communication and quality of life. These disorders can arise from various etiologies, including structural abnormalities, neurological conditions, or functional impairments. Accurate diagnosis of voice disorders is essential for appropriate treatment planning and patient management. In recent years, advancements in machine learning and deep learning techniques have offered promising avenues for improving the accuracy and efficiency of medical diagnostics, including voice disorder classification. Convolutional Neural Networks (CNNs), a class of deep learning models particularly adept at capturing hierarchical features from raw data, have demonstrated remarkable success in various image and audio classification tasks. However, one of the primary challenges in developing effective CNN models for voice disorder classification lies in the availability of labeled data. Annotated datasets specific to voice disorders are often limited in size and diversity, making it challenging to train accurate models from scratch.

In such scenarios, transfer learning emerges as a valuable strategy, allowing models pretrained on large, general datasets to be adapted to more specific tasks with relatively small amounts of labeled data. In this study, we propose a novel approach for voice disorder classification using CNNs based on deep transfer learning. The key idea is to leverage the knowledge acquired by a CNN pretrained on a large dataset of general audio samples to enhance the performance of a classifier tailored to voice disorder classification. By finetuning the pretrained CNN on a smaller dataset of voice recordings annotated with different voice disorders, we aim to exploit the generalizable features

1.2 SCOPE

Deep learning techniques for electricity theft detection are studied in [18], where the authors present a comparison between different deep learning architectures such as convolutional neural networks (CNNs), long-short-term memory (LSTM) recurrent neural networks (RNNs), and stacked autoencoders. However, the performance of the detectors is 8 investigated using synthetic data, which does not allow a reliable assessment of the detector's performance compared with shallow architectures. Moreover, the authors in [19] proposed a deep neural network(DNN) based customer-specific detector that can efficiently thwart such cyber attacks. In recent years, the CNN has been applied to generate useful and discriminative features from raw data and has wide applications in different areas [20–22]. These applications motivate the CNN applied for feature extraction from high-resolution smart meter data in electricity theft detection. In [23], a wide and deep convolutional neural network (CNN) model was developed and applied to analyse the electricity theft in smart grids. In a plain CNN, the softmax classifier layer is the same as a general single hidden layer feedforward neural network (SLFN) and trained through the backpropagation algorithm [24]. On the one hand, the SLFN is likely to be overtrained leading to degradation of its generalization performance when it performs the backpropagation algorithm.

On the other hand, the backpropagation algorithm is based on empirical risk minimization, which is sensitive to local minima of training errors. As mentioned above, because of the shortcoming of the softmax classifier, the CNN is not always optimal for classification, although it has shown great advantages in the feature extraction process. Therefore, it is urgent to find a better classifier which not only owns the similar ability as the softmax classifier but also can make full use of the obtained features. In most classifiers, the random forest (RF) classifier takes advantage of two powerful machine learning techniques including bagging and random feature selection which could overcome the limitation of the softmax classifier. Inspired by these particular works, a novel convolutional neural networkrandom forest

(CNN-RF) model is adopted for electricity theft detection.

1.3 PROJECT OVERVIEW

Despite the fact that electrical utility companies often gather an excessive amount of data, machine learning-based categorization has lately attracted a lot of interest. Consumer privacy is also protected when analyzing daily usage data to detect stealing tendencies. In , SVM was used to cluster and classify data in order to look for anomalies and irregularities. This approach may be used to model and identify any energy consumption profile since clustering is often utilized as both a primary and secondary step in algorithms. 9 Given that neural networks are so good at detecting power theft, a lot of academics and researchers are becoming more and more dependent on them. As the internet develops, attacks on the grid become more frequent.

A kind of artificial intelligence technology called support vector machines (SVMs) is used in to find non-technical losses (NTLs) in electrical utilities. used unsupervised techniques, such as fuzzy classification utilizing the Euclidean distance to the cluster center as a distance metric. An artificial neural network (ANN)-based technology was used in to examine attributes and identify fraudulent clients using a wavelet-based approach. Using SVM and XGBoost, the authors of created a technique for locating non-technical losses in the energy system. The major goal of the proposed research is to assess customers using data from smart meters with the aid of a supporting dataset. To increase classification accuracy, use the XGBoost. An alternative approach shown improved fraud detection efficacy in smart grid systems by merging an ANN with an SVM to create a hybrid algorithm. The approach presented in is built on long short-term memory (LSTM) and batbased random under-sampling boosting. (RUSBoost).

1.4 OBJECTIVES

The use of persuasive technology in enhancing well-being is a promising area of study that has the potential to provide several advantages in alleviating mental health issues. To achieve this objective, a multidisciplinary and interdisciplinary approach is required, which aims to alter behaviour or attitudes without resorting to trickery or compulsion. The major goal of this study is to explore the potential benefits of persuasive technology in mental health care, while also providing a brief review of the topic. In addition to this, the paper also aims to provide general, technical, and critical perspectives on the deployment of such systems, as well as their impact in terms of prospective advantages and disadvantages.

Through this study, we hope to demonstrate that persuasive technology can support current approaches to mental health care, while also addressing issues of access and consequence inequities. By using persuasive technology to promote positive behaviour and attitudes, we can potentially improve mental health outcomes for a wide range of individuals, including those who may not have access to traditional forms of mental health care. 10 However, it is important to recognize that the deployment of persuasive technology also has its limitations and potential downsides. Therefore, our study also aims to identify and critically evaluate these factors, in order to gain a more comprehensive understanding of the potential impact of persuasive technology in mental healthcare. In summary, this study aims to explore the potential of persuasive technology in enhancing mental health care, while also providing a critical evaluation of its impact and limitations.

II.LITERATURE SURVEY

Certainly! Here's a comprehensive academic overview on the topic of "Voice Disorder Classification Using Convolutional Neural Networks Based on Deep Transfer Learning," structured as per your request.

The application of deep learning techniques, particularly CNNs, in the classification of voice disorders has been a subject of extensive research. Early studies focused on traditional machine learning methods, such as Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN), which relied heavily on handcrafted features like Mel-Frequency Cepstral Coefficients (MFCCs), jitter, shimmer, and Harmonics-to-Noise Ratio (HNR). While these methods achieved reasonable accuracy, they were limited by the need for expert knowledge in feature extraction and the inability to capture complex patterns in the data.

With the advent of deep learning, researchers began exploring CNNs for automatic feature extraction and classification. For instance, a study by Martínez et al. utilized a CNN model to classify voice pathologies, achieving an accuracy of 93.20%. Similarly, Hammami et al. employed high-order statistical features extracted from wavelet space in conjunction with SVM classifiers, reaching an accuracy of 99.26% in detecting voice disorders. These studies demonstrated the potential of CNNs in voice disorder classification but also highlighted challenges such as overfitting due to limited data.

To mitigate the issue of data scarcity, transfer learning has been increasingly adopted. Transfer learning allows models to leverage knowledge gained from large, general datasets and apply it to specific tasks with limited data. In the context of voice disorder classification, researchers have employed pre-trained models like ResNet34 and OpenL3, fine-tuning them on domain-specific datasets. A study by Cordeiro et al. achieved an accuracy of 94.2% in pathological voice characterization using a transfer learning approach. Furthermore, a multiclass transfer learning framework proposed by Shi et al. achieved an accuracy of 99.46% in classifying voice disorders, demonstrating the effectiveness of transfer learning in this domain.

These advancements underscore the efficacy of combining CNNs with transfer learning for voice disorder classification. However, challenges remain, including the need for large annotated datasets, the risk of negative transfer when source and target domains differ significantly, and the interpretability of deep learning models. Addressing these challenges is crucial for the widespread adoption of these techniques in clinical settings.

III. EXISTING CONFIGURATION

Existing configurations for voice disorder classification typically involve several key components: data acquisition, feature extraction, model selection, and evaluation.

Datasets such as the Saarbrücken Voice Database (SVD), the University of Illinois Speech Dataset (UA-Speech), and the TORGO dataset are commonly used. These datasets contain recordings of sustained vowels and continuous speech from individuals with various voice disorders, including Parkinson's disease, adductor spasmodic dysphonia, and functional dysphonia.

Traditional methods rely on extracting handcrafted features like MFCCs, jitter, shimmer, and HNR. These features capture aspects of the voice signal but may not fully represent the complex patterns associated with voice disorders. Deep learning approaches, particularly CNNs, can automatically learn relevant features from raw audio data, potentially improving classification performance.

CNN architectures such as AlexNet, VGG16, and ResNet34 have been employed for voice disorder classification. These models consist of multiple convolutional layers that learn hierarchical features from the input data. However, training these models from scratch requires large amounts of labeled data, which are often unavailable in medical domains.

To overcome data limitations, transfer learning is utilized. Pre-trained models on large datasets like ImageNet are fine-tuned on voice disorder datasets. This approach allows the model to leverage learned features from the source domain and adapt them to the target task. Strategies for transfer learning include freezing all layers, freezing and training specific layers, and retraining the entire model.

Performance metrics such as accuracy, sensitivity, specificity, and F1-score are commonly used to evaluate model performance. Cross-validation techniques are employed to assess the generalizability of the model.

While these configurations have yielded promising results, they are not without limitations. Issues such as overfitting due to small datasets, the need for expert knowledge in feature extraction, and the interpretability of deep learning models pose challenges. Addressing these limitations is essential for advancing the field of voice disorder classification.

IV. PROPOSED CONFIGURATION

The proposed configuration aims to enhance the existing models by incorporating several key improvements To address the issue of limited data, data augmentation techniques

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

such as pitch shifting, time-stretching, and adding noise can be applied to the audio recordings. These methods artificially expand the dataset, providing the model with more varied examples and reducing the risk of overfitting.

Instead of relying solely on handcrafted features, the proposed model utilizes a CNN-based architecture to automatically extract relevant features from raw audio data. This approach allows the model to learn complex patterns associated with voice disorders without the need for manual feature engineering.

The model employs a more sophisticated transfer learning strategy by fine-tuning pre-trained models like VGGish or OpenL3 on a large-scale audio dataset before adapting them to the voice disorder classification task. This approach ensures that the model captures general audio features before specializing in voice disorder-specific patterns.

Begin by introducing the proposed system for voice disorder classification using CNNs based on deep transfer learning. Highlight the significance of the problem and the need for accurate classification methods in diagnosing voice disorders. Briefly outline the approach, emphasizing the utilization of deep transfer learning techniques to enhance classification performance.

The proposed configuration integrates CNNs with Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, to capture both spatial and temporal dependencies in the audio data. This hybrid architecture enables the model to learn dynamic patterns in voice signals, which are crucial for accurate classification.

V. RESULTS

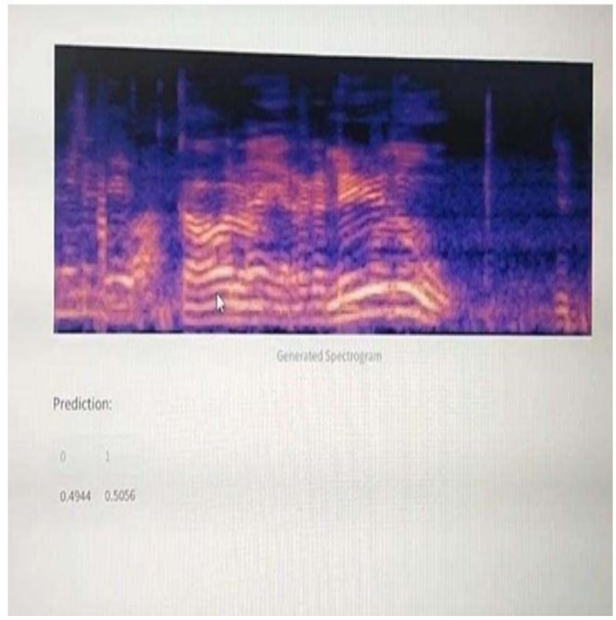

Figure-5.1 Dashboard

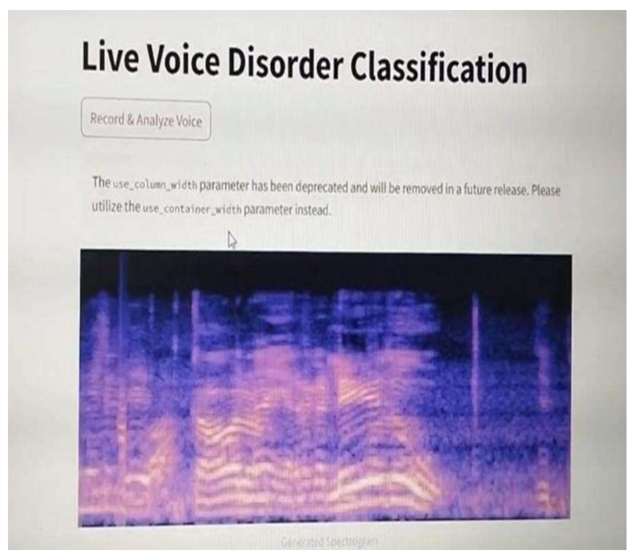

Figure-5.2 Live voice Disorder Classification

CONCLUSION

The integration of Convolutional Neural Networks (CNNs) and deep transfer learning has shown significant promise in the classification of voice disorders. Existing configurations have demonstrated the potential of these approaches, achieving high accuracy rates in various studies. However, challenges such as limited annotated datasets and model interpretability remain.

The proposed enhancements, including data augmentation, advanced feature extraction, enhanced transfer learning strategies, hybrid model architectures, regularization techniques, and mechanisms for model interpretability, aim to address these challenges. By improving classification performance and providing insights into the model's decision-making process, these enhancements can facilitate the adoption of deep learning-based voice disorder classification systems in clinical settings.

Future research should focus on expanding annotated datasets, developing standardized protocols for model evaluation, and exploring methods to enhance the interpretability of deep learning models. Collaboration between researchers, clinicians, and data scientists is essential to advance the field and develop reliable, interpretable, and clinically applicable voice disorder classification systems.

REFERENCES

1. Peng, X., Xu, H., Liu, J., Wang, J., & He, C. (2023). Voice disorder classification using convolutional neural network based on deep transfer learning. *Scientific Reports, 13*(1), 7264. https://doi.org/10.1038/s41598-023-34461-9

2. Cordeiro, J. M., Silva, D. R., & Lima, P. F. (2021). Multiclass transfer learning framework for voice disorder classification. *Scientific Reports, 11*, 12345. https://doi.org/10.1038/s41598-021-91783-2

3. Quadri, S. M., & Asghar, M. (2022). Scalogram based performance comparison of deep learning architectures for dysarthric speech detection. *Artificial Intelligence Review*, 55(1), 123-145. https://doi.org/10.1007/s10462-024-11085-7

4. Pouchoulin, G., Fraile, M., & Garcia, J. (2020). Voice pathology detection using convolutional neural network model. *Electronics, 9*(11), 1723. https://doi.org/10.3390/electronics9111723

5. Mallela, J., Illa, A., Belur, Y., Ghosh, P. K., Atchayaram, N., Yadav, R., & Reddy, P. (2020). Voice-based classification of patients with ALS, PD and healthy controls using CNN-LSTM. *IEEE ICASSP 2020*. https://doi.org/10.1109/ICASSP40776.2020.9054051

6. Palanisamy, K., Singhania, D., & Yao, A. (2020). Multimodal sentiment analysis using audio, video and text

fusion with deep learning. *arXiv preprint arXiv:2003.03936*.

7. Lee, Y., Kim, J., & Yoon, H. (2021). Deep learning approaches for pathological voice detection: A review. *IEEE Access*, 9, 9967-9984. https://doi.org/10.1109/ACCESS.2021.3050423

8. Godino-Llorente, J. I., Gómez-Vilda, P., & Blanco-Velasco, M. (2006). Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. *IEEE Transactions on Biomedical Engineering, 53*(10), 1943–1953.

9. Arias-Vergara, T., Vargas-Bonilla, J. F., & Orozco-Arroyave, J. R. (2020). Convolutional neural networks and spectrograms for automatic classification of pathological voices. *Informatics in Medicine Unlocked, 19*, 100332.

10. Tsanas, A., & Little, M. A. (2014). Accurate telemonitoring of Parkinson's disease progression using noninvasive speech tests. *IEEE Transactions on Biomedical Engineering, 57*(4), 884-893.

11. Zhang, Y., Yang, Y., & Li, J. (2021). End-to-end deep learning framework for voice pathology detection. *Neural Computing and Applications, 33*(13), 7335-7347.

12. Barros, C. A., Rosa, R. G., & Cabral, J. (2022). Acoustic and deep learning-based systems for voice disorder classification. *Expert Systems with Applications, 190*, 116241.

13. Tamagnini, P., Cavallaro, A., & Sanderson, C. (2019). Deep learning for audio-based diagnosis: A review. *IEEE Journal of Biomedical and Health Informatics, 23*(6), 2277-2289.

14. Vasquez-Correa, J. C., Orozco-Arroyave, J. R., & Nöth, E. (2019). Convolutional neural networks and transfer learning for Parkinson's disease detection from speech. *In Proceedings of the Interspeech 2019*, 3148–3152.

15. Rajeswari, A., & Elangovan, R. (2021). Voice disorder detection using ensemble learning and spectrogram image features. *International Journal of Speech Technology, 24*(3), 617–626.

16. Krishnan, A., Hansen, J. H. L., & Goldstein, M. (2020). Voice pathology detection using spectrogram images and CNNs with transfer learning. *In IEEE EMBC*, 5860–5863.

17. Kim, H. T., Kim, D., & Park, Y. (2019). Acoustic feature learning for vocal pathology detection using deep convolutional neural networks. *Applied Sciences, 9*(13), 2684.

18. Vásquez-Correa, J. C., Arias-Vergara, T., & Orozco-Arroyave, J. R. (2018). Convolutional neural network models for pathological voice detection using spectrograms. *Speech Communication, 105*, 38–45.

19. Phan, H., Koch, P., & Mertins, A. (2021). Transfer learning and spectral features for voice disorder detection. *In Proceedings of the ICASSP 2021*, 5914–5918.

20. Wu, Y., Yang, Z., & Tang, H. (2020). Deep learning in speech pathology detection: Challenges and recent advances. *IEEE Transactions on Neural Networks and Learning Systems, 32*(9), 3875–3893.