International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Pathology Classification in Voice Signal

Kartikeya Hegde[1], Karthik KS[2], Karthik SR[3]

[1]*Department of Computer Science And Engineering MITE Moodabidri* 574225.
[2]*Department of Computer Science And Engineering MITE Moodabidri* 574225.
[3]*Department of Computer Science And Engineering MITE Moodabidri* 574225.

---***---

ABSTRACT - Voice pathology classification is the process of identifying different types of voice disorders from voice signals. With the advancement of machine learning techniques, researchers have explored the use of automated systems for the classification of different voice disorders. The process of pathology classification in voice signals involves several steps, including signal acquisition, feature extraction, and classification. The signal acquisition step involves recording the voice signal of the patient. The feature extraction step involves extracting relevant features from the voice signal. These features are then used to train a classification model using machine learning algorithms. From the voice datasets, create a Mel spectrogram for a particular voice from the dataset. The proposed approach employs a convolutional neural network (CNN) trained on a dataset of voices with three corresponding labels, which are Laryngol, normal, and vox senilis. The experiments show that MobileNetV2 achieves high accuracy compared to other models. The paper also discusses the performance of MobileNetV2, such as the number of layers and the learning rate. The accuracy of the pathology classification model depends on several factors, such as the quality of the voice signal, the choice of features, and the algorithm used for classification.

Key Words**:** Voice pathology**,** Signal acquisition**,** Mel spectrogram**,** three labels**,** MobileNetV2

1. INTRODUCTION

The study of pathology categorization in voice signals focuses on examining vocal traits to find and identify different diseases or disorders. Insights concerning a patient's health status can be gained through voice signal analysis because vocal alterations can be early signs of conditions including Parkinson's, Alzheimer's, and other cancers. In order to categorize voice signals into distinct groups based on particular characteristics, such as pitch, frequency, and volume, machine learning methods are used for pathology classification in audio signals. The ability of this technology to diagnose patients more quickly and accurately has the potential to change the medical sector.

2. OBJECTIVE

The objective of pathology classification in voice signal analysis is to analyze the voice from a dataset and create a spectrogram image. By using the MobileNetV2 model to extract the features from the spectrogram image, it is possible to classify voice signals into different categories, such as normal, ox senilis, or diseased. The primary goal of healthy and non-healthy categorization in the sound signal analysis is to increase result accuracy through the use of various models such as ResNet, Inception V2, GoogleNet, and AlexNet.

3. LITERATURE REVIEW

"Automatic classification of voice disorders from acoustic measurements", by Maryn et al. (2011) In this study, the authors proposed an automatic system for the classification of different voice disorders based on acoustic measurements. They used a dataset of 504 voice samples from patients with vocal nodules, polyps, and other disorders. The system used machine learning algorithms, including support vector machines (SVM) and neural networks (NN), to classify the voice samples. The results showed that the SVM model achieved an accuracy of 91.1%, while the NN model achieved an accuracy of 88.9%.

"Classification of pathological voices using a hybrid approach based on neural networks and decision trees" by Gómez-Vilda et al. (2005) this study proposed a hybrid approach for the classification of pathological voices. The system used a combination of neural networks and decision trees to classify the voice signals. The dataset consisted of 234 voice samples from patients with different vocal disorders. The results showed that the system achieved an overall accuracy of 92.31%.

"Automatic classification of voice disorders using support vector machines" by Bozkurt et al. (2008) In this study, the authors proposed an automated system for the classification of voice disorders using support vector machines (SVM). The system used a dataset of 350 voice samples from patients with different vocal disorders. The results showed that the SVM model achieved an accuracy of 87.7% in classifying the different disorders.

"Classification of voice disorders using Mel frequency cepstral coefficients and a radial basis function neural network" by Bacanl et al. (2016) In this study, the authors proposed a system for the classification of different voice disorders using mel frequency cepstral coefficients (MFCC) and a radial basis function neural network (RBFNN). The dataset consisted of 230 voice samples from patients with different vocal disorders. The results showed that the system achieved an overall accuracy of 92.17%.

4. METHODOLOGY

The main goals of the project are to identify features that improve the identification and categorization of voice pathology as well as to look into how processes in various frequency regions (bands) affect these processes. The audio signals are processed before being fed into a convolutional neural network (CNN). We employ a transfer-learning framework and use stable existing CNN models.

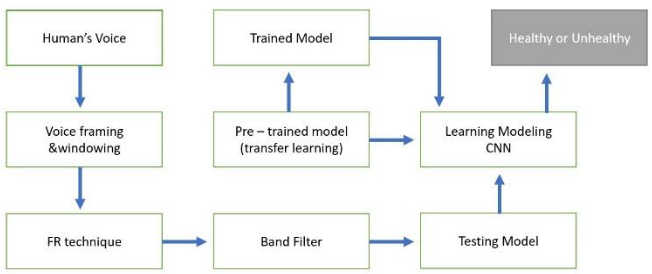
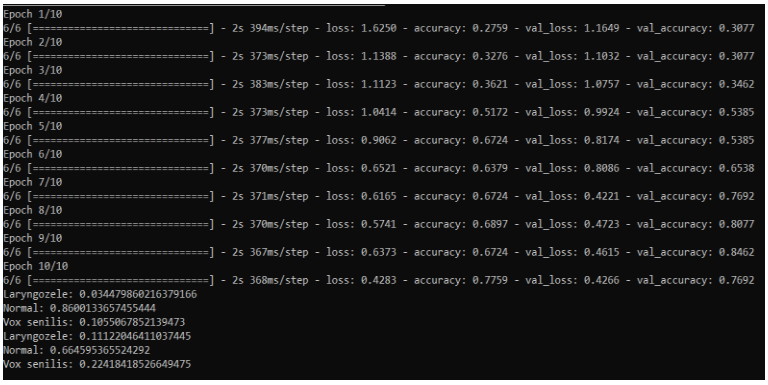


Fig -1: Work flow

The proposed deep learning method for speech pathology identification is shown in Figure 1. When a patient's voice is supplied into the system, the system's output determines whether the patient's voice is pathological or normal. The voice signal lasts one second. A central signal of 1 s is turned off when an input reaches 1 s. The signal is divided into 40-ms frames, with a 20-ms break between each frame. A well-balanced pitch capture and voice break smoothing choice is the 40-ms frame duration. If this goes on for a very long time, the vocal folds may open and close erratically due to voice breaks or other sounds. If the frame length is too short, the continuation effect and pitch duration are lost.
A quick Fourier transformation turns the framed signal into a frequency-domain signal. After joining all of the frames' frequency domains, we obtain a spectrogram. One could consider the spectrogram to be an image. There are at least 20 band pass filters in the spectrogram. The octave is the basis for the filters. around there The octave scale often performs better than the Mel scale for voice pathology identification [34]. For the octave spectrum output, first and second order time derivatives are used. Following this procedure, we obtain three patterns that resemble images: the octave and its first and second order derivatives. Three picture patterns make up the CNN model's input. ResNet34 was examined using the suggested technique.

5. RESULTS

Classified output three labels: laryngozele, normal, and vox senilis. Output in the form of a decimal.



```
Epoch 1/10
6/6 [==============================] - 2s 394ms/step - loss: 1.6250 - accuracy: 0.2759 - val_loss: 1.1649 - val_accuracy: 0.3077
Epoch 2/10
6/6 [==============================] - 2s 373ms/step - loss: 1.1388 - accuracy: 0.3276 - val_loss: 1.1032 - val_accuracy: 0.3077
Epoch 3/10
6/6 [==============================] - 2s 383ms/step - loss: 1.1123 - accuracy: 0.3621 - val_loss: 1.0757 - val_accuracy: 0.3462
Epoch 4/10
6/6 [==============================] - 2s 373ms/step - loss: 1.0414 - accuracy: 0.5172 - val_loss: 0.9924 - val_accuracy: 0.5385
Epoch 5/10
6/6 [==============================] - 2s 377ms/step - loss: 0.9062 - accuracy: 0.6724 - val_loss: 0.8174 - val_accuracy: 0.5385
Epoch 6/10
6/6 [==============================] - 2s 370ms/step - loss: 0.6521 - accuracy: 0.6379 - val_loss: 0.8086 - val_accuracy: 0.6538
Epoch 7/10
6/6 [==============================] - 2s 371ms/step - loss: 0.6165 - accuracy: 0.6724 - val_loss: 0.4221 - val_accuracy: 0.7692
Epoch 8/10
6/6 [==============================] - 2s 370ms/step - loss: 0.5741 - accuracy: 0.6897 - val_loss: 0.4723 - val_accuracy: 0.8077
Epoch 9/10
6/6 [==============================] - 2s 367ms/step - loss: 0.6373 - accuracy: 0.6724 - val_loss: 0.4615 - val_accuracy: 0.8462
Epoch 10/10
6/6 [==============================] - 2s 368ms/step - loss: 0.4283 - accuracy: 0.7759 - val_loss: 0.4266 - val_accuracy: 0.7692
Laryngozele: 0.034479860216379166
Normal: 0.8600133657455444
Vox senilis: 0.1055067852139473
Laryngozele: 0.11122046411037445
Normal: 0.664595365524292
Vox senilis: 0.22418418526649475
```

3. CONCLUSIONS

The ability to better identify and treat vocal problems early on has made pathology categorization in voice signals an important and active topic of research in recent years. In developing automated systems for the detection and categorization of vocal disorders, great progress has been made thanks to improvements in digital signal processing techniques and machine learning algorithms. It has been discovered that the analysis of voice signals is a reliable and non-invasive method for diagnosing a variety of vocal diseases, such as dysphonia, laryngitis, and vocal fold nodules. Machine learning algorithms can be trained to reliably classify various vocal illnesses by extracting pertinent information from voice signals, such as fundamental frequency, jitter, and shimmer.

However, there are still a number of difficulties in creating reliable and accurate pathology classification systems. These include the individualised variations in voice signals, the presence of noise in real-world settings, and the requirement for sizable datasets with top-notch annotations for developing machine learning models.

Despite these difficulties, pathology classification in voice signals holds great promise for enhancing vocal disorder diagnosis and care. It is anticipated that automated voice analysis systems will become more dependable and available in the near future as a result of ongoing research and technology breakthroughs, which will enhance the outcomes for patients with vocal disorders.

REFERENCES

[1] Mazin Abed Mohammed College of Computer Science and Information Technology, University of Anbar, 11, Ramadi 31001, Anbar, Iraq.

[2] Karrar Hameed Abdulkareem College of Agriculture, Al-Muthanna University, Samawah 66001, Iraq; khak9784@mu.edu.iq

[3] Salama A. Mostafa, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Batu Pahat 86400, Malaysia; salama@uthm.edu.my

[4] Mohd Khanapi Abd Ghani Biomedical Computing and Engineering Technologies (BIOCORE) Applied Research Group, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Durian Tunggal 76100, Malaysia; khanapi@utem.edu.my

[5] Mashael S. Maashi Software Engineering Department, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia; mmaashi@ksu.edu.sa

[6] Begonya Garcia-Zapirain Ibon Oleagordia eVIDA Lab., University of Deusto, Avda/Universidades 24, 48007 Bilbao, Spain; mbgarciazapi@deusto.es (B.G.-Z.); ibruiz@deusto.es (I.O.)

[7] Hosam Alhakami, Department of Computer Science, College of Computer and Information Systems, Umm Al-Qura University, Makkah 21421, Saudi Arabia; hhhakam@uqu.edu.sa

[8] Fahad Taha AL-Dhief, Faculty of Engineering, School of Electrical Engineering, Universiti Teknologi Malaysia (UTM), Johor Bahru 81310, Malaysia; taha-1989@graduate.utm.my.