Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230721

A deep learning assisted web application tool for diagnosing communication disorder in children

C. S. KanimozhiSelvi[1], S. Santhiya[1], Sharmila C.[2], Gayathri S.[2], Venkatesh V.[2] and Snehaa S.[2]

[1]Department of Artificial Intelligence, Kongu Engineering College, Perundurai, India
[2]Department of CSE, Kongu Engineering College, Perundurai, India

santhiya123cse@gmail.com

Abstract. Communication is an important way of expressing one's thoughts to others. Many people are suffering from communication disorders like stammering and stuttering. Communicative disorders are mainly divided into three types namely language disorders, speech production disorders, and oral motor/swallowing/feeding disorders. Identifying the problem earlier and giving intervention at the right time is very important for children to improve their personal and academic lives. It will be better for the children to recover from the disorder if the problem is identified earlier. Disorder identification is a complex decision-making process that requires expertise in the medical field. The proposed framework gets the voice input through a web application. Later, deep learning technology predicts the speech disorder of the recorded voice. The deep learning-assisted tool developed in the proposed framework is to identify if the person has a stuttering problem, dysarthria, or whether the person can talk normally. Hence, the proposed method alleviates the problem of identifying speech disorder types by recording the person's voice through a web application and it achieves the accuracy of 95%.

Keywords: machine learning, deep learning, web application, communication disorder, convolutional neural network.

1. Introduction

Communication is a person's ability to share individual thoughts with others or a group of persons in a proper way. Speech difficulties are the most prevalent problem experienced by many people. Children with disorders are unable to participate in daily life activities which causes emotional and behavioral issues. A speech problem affects the person's capacity to produce sounds that form words. Various speech disorders are stuttering, apraxia, and dysarthria. Stuttering is a speech impairment distinguished by the repetition of sounds, syllables, or words, as also sound prolonging and speech disruptions known as blocks. Slurred or slow speeches are difficult to understand is a symptom of dysarthria. Speech-language disorders are the most prevalent childhood impairments, as they are also the least recognized, especially in the primary care stage. The proposed framework is to identify speech disorders using a deep learning-based web application. The web application is easy to use and available for free which makes it easy and efficient. Deep Learning is meant to contribute to major performance in the field of

DOI: 10.54254/2755-2721/6/20230721

medicine. It can process a large number of audio datasets beyond human capability and convert the analyzed data into clinical insights.

Identification of disorders using deep learning techniques at an early stage can be helpful for children to overcome several problems. Datasets are gathered and trained to create the model for predicting the disorder. In deep learning, a Convolutional Neural Network (ConvNet/CNN) takes an input image and gives priority to unique image properties to separate them. A ConvNet requires substantially less pre-processing when compared to other classification algorithms. The existing approaches require hand-engineering of filters but ConvNets learns the filters/characteristics during training for a large number of epochs.

2. Literature review

Level assessments have been developed for foreign language students. The project involves developing an audio processing system that uses five distinct machine learning models to classify the level of fluency of non-native English speakers [1]. A method has been proposed for the early identification of speech toddlers [2]. DCD is a motor skill impairment that affects a child's ability to complete age-appropriate self-care and academic responsibilities. Later, a Software system has been created that allows Persian Hearing-impaired youngsters to participate in and synchronize language learning exercises [3]. A software system had been developed for Persian children with articulation difficulties in response to flaws in the conventional speech therapy procedure. Moreover, an e-learning system has been developed that helps Greek preschoolers to improve their articulation [4]. ICTs have aided therapists in many situations, but in others, particularly in rural area, they have acted as the only providers of therapy.

The importance of early detection has been emphasized to help the students overcome the specific learning disorder [5]. The work was focused on the many methods and techniques for screening students with specific learning disorders while emphasizing the necessity of early detection of individuals with specific learning disorders. Children with autism were given computer-based therapies to help them develop their social skills. According to the results of numerous research, computer-based games are frequently utilised to help ASD children with their social skills [6].

Augmentative and Alternative Communication (AAC) prototype termed "ISpeak" was developed and focuses on all communication modes [7]. The computer-based intervention to improve communicative capabilities in autistic children was investigated and created a software program [8]. The software application has been created using real-life activities like play, food, and cleanliness. Moreover, analysis has been performed on audio datasets [9], which have been collected from children with and without defects using the Android app. It was observed that CNN outperformed better than computer-based education which aided in alleviating the challenges of identifying certain types of communication disorders.

A systematic review of studies was carried out on computer-based interventions (CBI) for teaching communication skills to children with autism spectrum disorders (ASD) [10]. The analyses included in the evaluation are intervention outcomes, assessing the certainty of the evidence, and specifying the software and system requirements. Videos of young children have been collected at the time of watching movies intended at eliciting autism-related behaviors. The outcomes of the Autism & beyond iPhone investigation were given which were based on the open-source ResearchKit framework [11].

Co-occurring mental health issues were described along with communication challenges [12]. Significant difficulties exist in education and public health. Confidence New diagnostic categories of social communication disorder (SCD) were specified which is part of the Diagnostic and Statistical Manual of Mental Disorders under Communication Disorders. Hence SCD shouldn't be diagnosed until they are 4–5 years old. Emotion recognition system has been focused on describing the accuracy of human emotion recognition. SLT designed and utilized the instrument to gather information about children and their emotions (self-emotion evaluation) [13].

DOI: 10.54254/2755-2721/6/20230721

3. Problem statement

A speech disorder is identified by speech-language pathologists. Speech-language pathologists are referred to as speech therapists. Therapists aid people of all ages with evaluation, diagnosis, and rehabilitation by using a variety of approaches and technologies. People can use the tools only with help of a Speech therapist for identifying the disorder. There is no proper therapy tool to particularly focus on fluency improving. The tools are cost-effective for common people in diagnosing the disorder. Early identification can be easily done using the proposed approach.

4. Proposed work

The research aims to predict the disorder based on the degree of accuracy in a speech signal. Deep learning techniques have been applied to classify the disorder based on the percentage obtained from the model. Figure 1 shows the workflow of an application.

4.1 Data collection

Voice recordings of people are collected for the dataset. Recordings consist of people saying a small sentence in Tamil language. The dataset is generally of different data formats. Dataset is converted into .json format. Each collected audio data has been labeled. 1955 audio recordings were collected from website research. google to load training dataset in .json format. It displays a different sentence for each time the website is loaded.

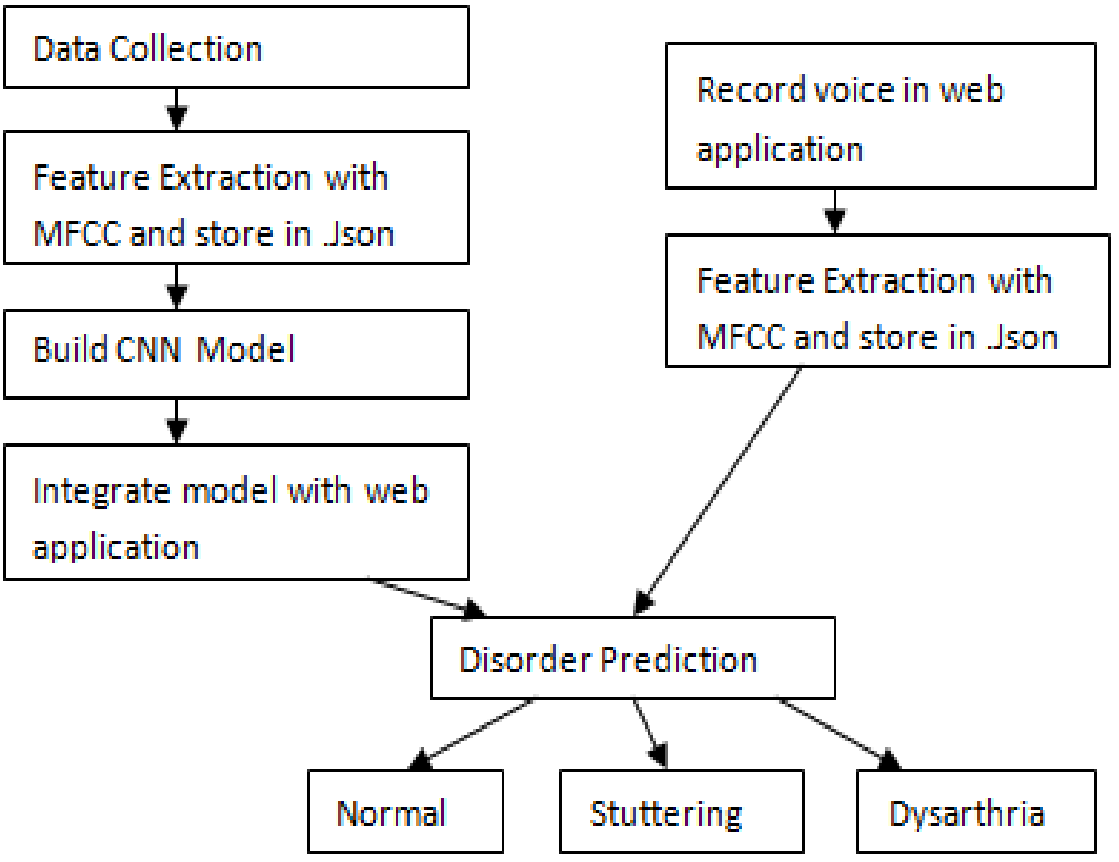


Figure 1. Workflow of the application.

DOI: 10.54254/2755-2721/6/20230721

4.2 Feature extraction

Feature extraction is a kind of dimensionality reduction in which it extract the features by employing Mel-frequency cepstral coefficients (MFCC). The feature extraction technique using MFCC includes windowing the signal, applying the DFT, taking the log of the magnitude, and then warping the frequencies on a Mel scale, followed by utilizing the inverse DCT. They are made up of the audio clip's cepstral representation. The MFCCs signal is a small set of features (usually 10-20) that define the overall shape of a spectral envelope. MFCC can be used as a feature for speech recognition. Audacity is used to sample and store the recorded voice signals. 16000 samples per second are used as the sampling rate. Each voice signal is separated into 16 ms windows, resulting in 256 samples.

4.3 Convolutional neural network

CNN works well for images. The hidden layers extract features by doing various calculations and manipulations. The image's pixels are passed into the convolutional layer, which performs the convolution and produces a convolved map. The convolved map is passed to a ReLU function to obtain a corrected feature map. Multiple convolutions and ReLU layers are used to locate the features in the image. Different pooling layers with varied filters are utilized to identify certain features. The pooled feature map is flattened and given to a fully connected layer to predict the final performance where it identifies the type of the disorder

4.4 Website development

The main objective of the website is to bring an interactive application to the user. The welcome screen prompts the user to select the preferred language for recording the voice input from the user. The page has a record button, a pause button, and a stop button. The record button needs to be clicked when the website is loaded and then the user can read the Tamil sentence which will be displayed on the screen. Once it is done, the recording will be displayed on the webpage. Save to disk and upload buttons will be displayed and enabled along with the recording. The recording will get saved into the downloads folder in the personnel computer if the save to disk button is clicked. If the upload button is clicked then the recording is stored in a folder inside the running program, on the local disk. Once the recording is uploaded and ready for prediction, click the predict button on the webpage. When the predict button is pressed, it calls a function. Then the audio is passed to the MFCC function and converted to a .json file. The model runs once the predict button is pressed. The file stored in the local disk is passed to the model and disorder is predicted.

4.5 Experiment results and discussion

Table 1. Performance metrics.

Class	Precision	Recall	F1-Score
Dysarthria	0.99	1.0	0.99
Normal	0.94	0.91	0.93
Stuttering	0.91	0.94	0.99

The website is a speech disorder recognition application which uses the person’s voice recording to identify the type of disorder. Accuracy is a metric for evaluating classification models. Mathematically, accuracy may be defined as the ratio of the number of correct predictions to the total number of predictions. Table 1. Shows the Performance Metrics of CNN Model.

The model yields a quite fair percentage (95%) of accuracy which reveals that most of the predictions for the type of disorder are correctly shown. Figure 2 shows the training accuracy and epoch of the CNN model.

DOI: 10.54254/2755-2721/6/20230721

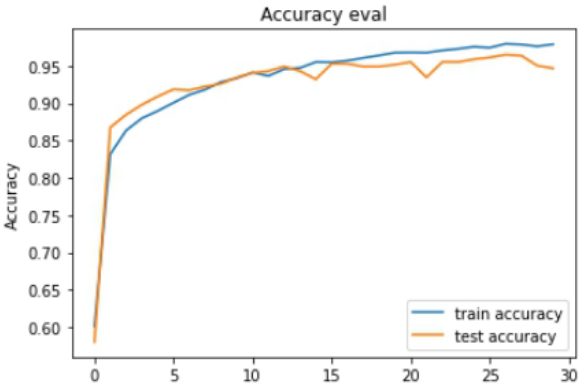


Figure 2. Training accuracy and epoch of the CNN model.

The website can predict disorders using simple recorded sentences depending on the user's speech by using a voice classification model. Figure 3,4,5 shows the GUI of the website. The user can record a sentence and gets the prediction of disorder type immediately once the website is opened.

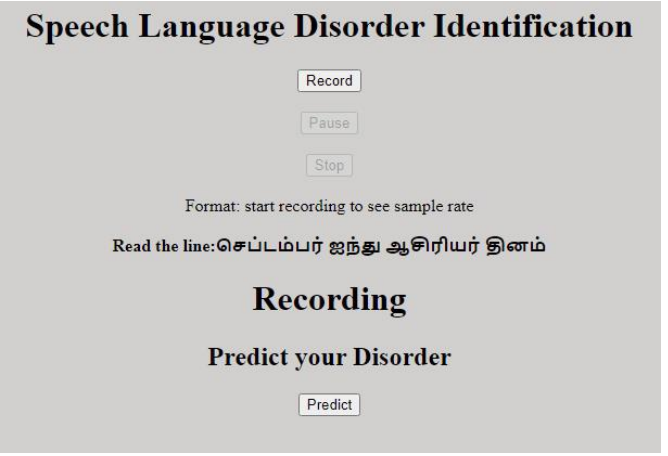


Figure 3. Website GUI - audio record.

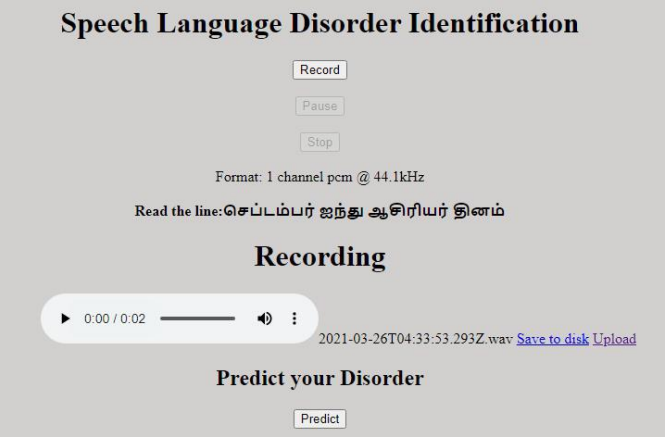
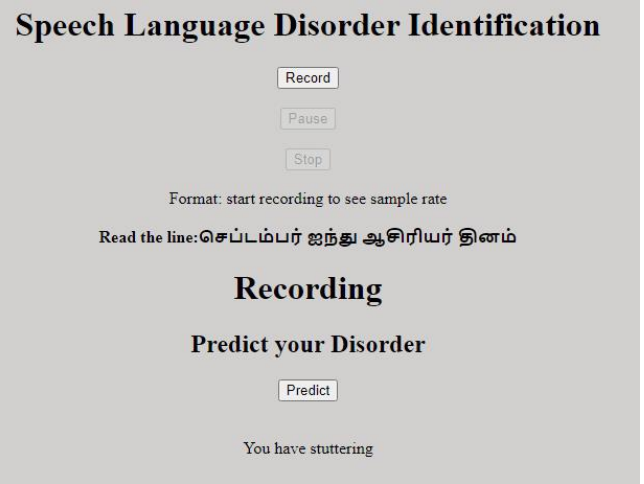


Figure 4. Website GUI - audio upload and save.

DOI: 10.54254/2755-2721/6/20230721



. **Figure 5.** Website GUI – Prediction.

5. Conclusion

The web application was developed for the welfare of people. Dataset consists of recordings of people saying a few sentences in the Tamil Language. The dataset consists of both normal and disordered data. Validation/Test data is given finally to check the accuracy and loss of the constructed model. The accuracy is found to be 95%. It will prompt the person in early identification of the disorder. So treating as early as possible can be done at a low cost. In near future, the work will be further extended by developing many more languages and many sentences.

Acknowledgment

This work is carried out as part of the Indian Council of Social Science Research (Impress) (ICSSR) Impactful Policy Research in Social Science sponsored project work titled Study, Design and Development of a Machine Learning-based Diagnostic and Therapeutic Application for Communicative Disorders

References

[1] Witsawakiti, N., Suchato, A., & Punyabukkana, P. (2006). Thai language e-training for the hard of hearing. Special Issue of the International Journal of the Computer, the Internet and Management, 14(SP1).

[2] Konstantinidis, E. I., Hitoglou-Antoniadou, M., Luneski, A., Bamidis, P. D., & Nikolaidou, M. M. (2009). Using effective avatars and rich multimedia content for the education of children with autism. In Proceedings of the 2nd international conference on pervasive technologies related to assistive environments (pp. 1-6).

[3] Zhang, M., Wang, X., Sathishkumar VE., & Sivakumar, V (2021). Machine learning techniques based on security management in smart cities using robots. Work, 68(3), 891-902.

[4] E.I. Toki and J. Pange, (2010) "E-learning activities for articulation in speech-language therapy and learning for preschool children," Procedia-Social and Behavioral Sciences, 2(2), 4274-4278.

[5] Kaur, A., & Padmanabhan, J. (2017). Children with specific learning disorder: identification and interventions. Educational Quest, 8(1), 1.

[6] Subramanian, M., Kumar, M. S., Sathishkumar, V. E., Prabhu, J., Karthick, A., Ganesh, S. S., & Meem, M. A. (2022). Diagnosis of retinal diseases based on Bayesian optimization deep

DOI: 10.54254/2755-2721/6/20230721

learning network using optical coherence tomography images. Computational Intelligence and Neuroscience, 2022.

[7] Pavithra, E., Janakiramaiah, B., Narasimha Prasad, L. V., Deepa, D., Jayapandian, N., & Sathishkumar, V. E. (2022). Visiting Indian Hospitals Before, During and After COVID. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.

[8] Hetzroni, O. E., & Tannous, J. (2004). Effects of a computer-based intervention program on the communicative functions of children with autism. Journal of autism and developmental disorders, 34(2), 95-113.

[9] Kanimozhiselvi, C. S., & Santhiya, S. (2021, February). Communication Disorder Identification from Recorded Speech using Machine Learning Assisted Mobile Application. In 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (pp. 789-793). IEEE.

[10] Ramdoss, S., Lang, R., Mulloy, A., Franco, J., O'Reilly, M., Didden, R., & Lancioni, G. (2011). Use of computer-based interventions to teach communication skills to children with autism spectrum disorders: A systematic review. Journal of Behavioral Education, 20(1), 55-76.

[11] Liu, Y., Sathishkumar, V. E., & Manickam, A. (2022). Augmented reality technology based on school physical education training. Computers and Electrical Engineering, 99, 107807.

[12] J. Pinborough-Zimmerman, R. Satterfield, J. Miller, D. Bilder, S. Hossain, and W. McMahon, [2007] "Communication disorders: Prevalence and comorbid intellectual disability, autism, and emotional/behavioral disorders," American Journal of Speech-Language Pathology

[13] Subramanian, M., Sathishkumar, V. E., Ramya, C., Kogilavani, S. V., & Deepti, R. (2022, May). A Lightweight Depthwise Separable Convolution Neural Network for Screening Covid-19 Infection from Chest CT and X-ray Images. In 2022 18th International Conference on Distributed Computing in Sensor Systems (DCOSS) (pp. 410-413). IEEE.