**Literature Review (Secondary Research) Template**

|  |  |
| --- | --- |
| **Student Name** | **Jasthi Pavithra** |
| **Project Topic Title** | **SpeechSentio: AI-powered Speech therapy with Emotion Analysis** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Type of Variables that You Need to Search for in Each Article (Each Current Solution)** | | | |
| **Dependent variable** | **Independent variable** | **Moderating variable** | **Mediating ( Intervening) variable** |
| * The presumed **effect** in an experimental study. * The values of those variable depend upon another variable that are the independent variables. * Strictly speaking, “dependent variable” should not be used when writing about non-experimental designs. | * The presumed **cause** in an experimental study. * The variables that may impact on the dependent variable * The values of those variable are under experimenter control. * Strictly speaking, “independent variable” should not be used when writing about non-experimental designs. | * has a strong  *contingent*effect on the independent variable-dependent variable **relationship** and thus produces an interaction effect. | * It comes between the independent and dependent variables and shows the **link or  mechanism** between them. |
| * Examples: **1.** **performance**. **2.** **Test Score**. **3.** **stock market. 4. performance** of the  students | * Examples: **1.** **run time** that will impact and cause high/low performance. **2.** **Time Spent Studying** that will cause the high/low score. **3.** **New product**  that will impact on the  stock market price. **4.** **quality of  library facilities** | * Example: **4.** There is a strong relationship  between the quality of  library facilities  (X) and the performance of the  students  (Y). Only  those students who have the **interest and  inclination** to use the  library  will show improved performance in their studies, which moderates the strength of the association between X and Y variables. | * Example: Parents transmit their social status to their children directly, but they also do so indirectly, through education: viz. Parent’s status ➛ child’s education ➛ child’s status * Example: The statistical association between income and longevity needs to be explained because just having money does not make one live longer. Other variables intervene between money and long life. People with high incomes tend to have better medical care than those with low incomes. Medical care is an intervening variable. It mediates the relation between income and longevity. |

|  |
| --- |
| **Relationship among Variables - Correlations (Univariate, Bivariate, Multivariate)** |
| * Once the variables relevant to the topic of research have been identified, then the researcher is interested in the relationship among them. * A statement containing the variable is called a **proposition**. It may contain one or more than one variable. * The proposition having one variable in it may be called as **univariate**  proposition,  those with two  variables as **bivariate**  proposition, and then of course  **multivariate** containing  three or more variables. * Prior to the formulation of a proposition the researcher has to develop strong  logical arguments  which could help  in establishing the  relationship. * For example, age at marriage and education are the two variables that could lead to a proposition: the higher the education, the higher the age at marriage .  What could be the logic to reach this conclusion? All relationships have to be explained with strong logical arguments. If the relationship refers to an observable  reality, then the proposition can be put to test, and any testable proposition is hypothesis. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Research Model That The Author Followed to Propose His Solution** | | | |
| **1. Where we are now** | **2. Where are we going** | **3. How do we get there** | **4. How do we know when we are finished** |
| * What the author has done in the area; The constructs that the literature examine * **What the problem is available** in this paper that has solved by the author * The purpose of that is to avoid pursing research which has already been undertaken | * What the author **objective** of the research is to gain a clearer understanding the relationships between variables * What is the goal of the paper * The purpose is to know what is the plan to do before he did the research | * How the author conducted the research; **How the problem has solved** * How he analysed the data generated by the research; A quantitative research design | * What is the value of this solution * A series of **recommendations** which flow from the data analysis have been made |

**NOTE: Please you need to use YOUR OWN WORDS in writing this template.**

**Your Literature Review Should be in Scope and MUST Address all Your Project's Questions**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Version 1.0 \_ Week 1** | | | | | | | |
| **1** |
| **Reference in APA format** | |  | | | | | |
| **URL of the Reference** | | **Authors Names and Emails** | | | | **Keywords in this Reference** | |
| <https://www.sciencedirect.com/science/article/pii/S1877050920318512> | | Hadhami Aouani  Yassine Ben Ayed | | | | Emotion recognition, Mel-frequency Cepstral Coefficient, Zero Crossing Rate, Teager Energy Operator, Harmonic to Noise Ratio, Support Vector Machine, Auto-encoder | |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc )** | | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | | | | **What are the components of it?** | |
| Support Vector Machine (SVM) | | The objective is to create a robust emotion recognition system using speech signals. The challenge is to automatically identify and classify human emotions. The solution combines features like MFCC, HNR, ZCR, and TEO, utilizing SVM for classification. Auto-encoders are introduced for feature dimension reduction to enhance SVM performance. | | | | An emotion recognition system based on speech signals in two-stage approach, namely feature extraction and classification engine. | |
| **The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process** | | | | | | | |
| The paper proposes a new framework for speech emotion recognition using one-dimensional deep convolutional neural networks (CNNs) with the combination of five different audio features as input data. The authors evaluate their model on three public datasets, RAVDESS, IEMOCAP, and EMO-DB, and achieve state-of-the-art results on all three datasets.   |  |  |  |  | | --- | --- | --- | --- | |  | **Process Steps** | **Advantage** | **Disadvantage (Limitation)** | | **1** | 1. Extract five different features from a sound file:    * Mel-frequency cepstral coefficients (MFCCs)    * Spectral centroid    * Spectral roll-off    * Zero-crossing rate    * Energy 2. And Feed the extracted features into a one-dimensional convolutional neural network (CNN) | * The use of deep CNNs, which are able to learn complex patterns in the data without the need for handcrafted feature engineering. | While deep learning models have achieved state-of-the-art results on some public datasets, their accuracy in real-world conditions can be lower. | | **2** | 1. Train the CNN to predict the emotion of the speaker and Evaluate the trained CNN on a held-out test set | The combination of multiple audio features, which provides more information to the model and helps to improve its performance | Deep learning models can be sensitive to noise and interference in the speech signal. This can degrade their performance, especially in noisy environments. | | | | | | | | |
| **Major Impact Factors in this Work** | | | | | | | |
| <Find all main factors and variables that are related to each solutions. Then find the relationship between factors. (Independent variable) causes a change in (Dependent Variable) and it isn't possible that (Dependent Variable) could cause a change in (Independent Variable).   |  |  |  |  | | --- | --- | --- | --- | | **Dependent Variable** | **Independent Variable** | **Moderating variable** | **Mediating (Intervening ) variable** | | The emotion | Features used for emotion recognition like MFCC, HNR, ZCR, TEO | Auto-encoder | SVM Classifier | | | | | | | | |
| |  | | --- | | **Relationship Among The Above 4 Variables in This article** | | The problem is accurately identifying emotions from speech. Features like MFCC, HNR, ZCR, and TEO are independent variables. Emotion identification is the dependent variable. Auto-encoder moderates by reducing feature dimensionality, and SVM classifies emotions, serving as a mediating variable. Together, they create a system for enhanced speech emotion recognition. | | | | | | | | |
| **Input and Output** | | | **Feature of This Solution** | | | | **Contribution & The Value of This Work** |
| |  |  | | --- | --- | | **Input** | **Output** | | a raw audio file, or features extracted from the audio signal | A prediction of the speaker's emotion. | | | | It uses a combination of five different spectral representations of the same sound file as input to a deep learning model. This combination of features allows the model to better capture the nuances of human speech and emotion. | | | | A new framework for speech emotion recognition using deep learning. The authors combine five audio features as input to a one-dimensional deep CNN, and their model outperforms the state-of-the-art on two datasets. |
| **Positive Impact of this Solution in This Project Domain** | | | | | **Negative Impact of this Solution in This Project Domain** | | |
| It has the potential to improve the user experience, educational outcomes, customer service quality, healthcare delivery, and marketing effectiveness. | | | | | The Speech emotion recognition with deep learning can be biased, invade privacy, and be used for manipulation or security breaches. | | |
| **Analyse This Work By Critical Thinking** | | | | **The Tools That Assessed this Work** | | | **What is the Structure of this Paper** |
| This work is good and promising, but it is important to be aware of its limitations and potential negative impacts. It is important to use speech emotion recognition responsibly and to mitigate the potential risks. | | | | Support Vector Machine | | | Abstract   1. Introduction 2. Related work 3. Methods 4. Experiments and results 5. Conclusion |
| **Diagram/Flowchart** | | | | | | | |
|  | | | | | | | |

**---End of Paper 1-**

|  |
| --- |
| **2** |
| **Reference in APA format** | |  | | | | | |
| **URL of the Reference** | | **Authors Names and Emails** | | | | **Keywords in this Reference** | |
| <https://ijrpr.com/uploads/V3ISSUE5/IJRPR4210.pdf> | | Husbaan I. Attar, Nilesh K. Kadole, Omkar G. Karanjekar, Devang R. Nagarkar, Prof. Sujeet | | | | Speech Emotion Recognition (SER), Convolutional Neural Networks (CNN),Emotion Classification,Human-Machine Interaction Deep Neural Networks (DNN) | |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc )** | | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | | | | **What are the components of it?** | |
| Convolutional Neural Networks | | The goal of this solution is to advance affective computing, facilitating effective human-computer interaction by proposing a Real-Time Speech Emotion Recognition System. It aims to overcome this limitation by developing a system that recognizes emotions in real-time from continuous speech. | | | | It consists of Voice Activity Detection to identify speech segments, Speech Segmentation for meaningful division, Signal Pre-Processing for conditioning audio, Feature Extraction extracting relevant speech features, Emotion Classification utilizing machine learning, and Statistics Analysis of Emotion Frequency for insight. | |
| **The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process** | | | | | | | |
| |  |  |  |  | | --- | --- | --- | --- | |  | **Process Steps** | **Advantage** | **Disadvantage (Limitation)** | | **1** | It involves gathering speech data, particularly from the RAVDESS dataset, and pre-processing it by extracting features like Mel Frequency Cepstral Coefficients (MFCCs) using LibROSA. | 1. The proposed speech emotion recognition system demonstrates a high accuracy rate, reaching 90% in experiments. | * The system's effectiveness may vary depending on the complexity and nuances of different emotional states. | | **2** | 1. A Convolutional Neural Network (CNN) architecture is employed for speech emotion recognition. | 1. The system's application in online learning environments proves beneficial by efficiently recognizing students' responses to the course material. | * While the system performs well with pre-recorded datasets, its reliance on such data may pose challenges when faced with diverse and dynamic real-world scenarios. | | **3** | 1. The trained model is rigorously evaluated using both pre-recorded datasets and real-time recordings featuring various emotion categories. |  |  | | | | | | | | |
| **Major Impact Factors in this Work** | | | | | | | |
| |  |  |  |  | | --- | --- | --- | --- | | **Dependent Variable** | **Independent Variable** | **Moderating variable** | **Mediating (Intervening ) variable** | | Stuttering Severity | Automation of Stuttering Recognition | Age | Workload of Speech Language Pathologists | | | | | | | | |
| |  | | --- | | **Relationship Among The Above 4 Variables in This article** | | This study explores the connection between stuttering severity and the automation of stuttering recognition , with age serving as a moderating factor. The workload of Speech Language Pathologists acts as a mediating variable, influencing the relationship between automation and stuttering severity. The study aims to discern how automated recognition, moderated by age and mediated by workload, correlates with the severity of stuttering. | | | | | | | | |
| **Input and Output** | | | **Feature of This Solution** | | | | **Contribution in This Work** |
| |  |  | | --- | --- | | **Input** | **Output** | | The input is speech | The output is emotion | | | | These features collectively contribute to a comprehensive representation of speech signals, enabling the automated system to effectively recognize and quantify the severity of stuttering. The system's distinctive Statistics Analysis of Emotion Frequency provides insights into emotional prevalence. | | | | * The automatic recognition of stuttering severity is a significant advancement in clinical practice.This system ensures objective measurements, facilitates early interventions, contributes to research endeavours, enhances patient experiences, and promotes public awareness. Moreover, it offers the prospect of cost-efficient healthcare practices through streamlined assessments. |
| **Positive Impact of this Solution in This Project Domain** | | | | | **Negative Impact of this Solution in This Project Domain** | | |
| It introduces a real-time emotion recognition system for continuous speech, elevating human-computer interaction with advanced features like efficient Voice Activity Detection and comprehensive Speech Segmentation. The use of a Convolutional Neural Network (CNN) for Emotion Classification ensures accurate and responsive results, enhancing the user experience with a nuanced and personalized touch. | | | | | The complexity of implementing advanced features, particularly the CNN, may require specialized skills and resources, potentially limiting accessibility for some organizations. Additionally, the system's performance could be constrained by the quality and diversity of the training data, posing a risk of reduced accuracy, especially for less-represented emotional states. | | |
| **Analyse This Work By Critical Thinking** | | | | **The Tools That Assessed this Work** | | | **What is the Structure of this Paper** |
| This approach has the potential to enhance user experience by providing nuanced and personalized interactions. However, the complexity of implementation, reliance on representative training data, and ethical considerations pose challenges. The system's success depends on addressing these concerns and staying abreast of evolving technological landscapes and competing solutions. | | | | A combination of statistical analysis tools, machine learning frameworks like Tensor Flow or PyTorch, and speech processing libraries such as LibROSA. | | | Abstract   1. Introduction 2. Problem Statement 3. Existing system 4. Existing system algorithm 5. System Implementation 6. Implementation Diagram   Conclusion |
| **Diagram/Flowchart** | | | | | | | |
|  | | | | | | | |

**--End of Paper 2--**

|  |
| --- |
| **3** |
| **Reference in APA format** | |  | | | | | |
| **URL of the Reference** | | **Authors Names and Emails** | | | | **Keywords in this Reference** | |
| <https://arxiv.org/ftp/arxiv/papers/2007/2007.08003.pdf> | | Dr. Mrs. Gresha Bhatia, Binoy Saha, Mansi Khamkar, Ashish Chandwani , Reshma Khot | | | | Stutter diagnosis, Stuttering therapy, Stutter measurement, Speech dysfluency, Mel-frequency Cepstral Coefficients (MFCC), CNN, Gated Recurrent Units (GRU), Support Vector Machine (SVM) | |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc )** | | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | | | | **What are the components of it?** | |
| The proposed technique is Support Vector Machine (SVM) and Gated Recurrent CNN (GRCNN) Models | | The main objective is to improve a person's speech fluency by accurately diagnosing stutter and suggesting appropriate training exercises for practice. | | | | The components are stutter assessment, therapy suggestion, Gated Recurrent CNN (GRCNN) models, SVM model, and a mobile application. | |
| **The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process** | | | | | | | |
| |  |  |  |  | | --- | --- | --- | --- | |  | **Process Steps** | **Advantage** | **Disadvantage (Limitation)** | |  | Stutter Assessment involves dataset preparation, data labeling, and feature extraction using MFCC features. | The Stutter Diagnosis and Therapy System Based on Deep Learning provides quantitative analysis of stuttering, allowing for accurate diagnosis and assessment of the severity and type of stutter. | The lack of a well-structured dataset for stuttered speech poses a challenge in training the models and may result in false positives due to background noise | |  | Separate GRCNN models are trained for detecting prolongation and repetition in speech audio. | The system recommends appropriate speech therapies based on the patient's performance and improvement, providing personalized treatment options. |  | |  | An SVM model with a polynomial kernel is trained to recommend suitable therapies |  |  | | | | | | | | |
| **Major Impact Factors in this Work** | | | | | | | |
| |  |  |  |  | | --- | --- | --- | --- | | **Dependent Variable** | **Independent Variable** | **Moderating variable** | **Mediating (Intervening ) variable** | | Effectiveness of speech therapies | Stutter descriptors | Baseline speech proficiency | correlation between stutter descriptors and effectiveness of speech therapies | | | | | | | | |
| |  | | --- | | **Relationship Among The Above 4 Variables in This article** | | This study investigates how stutter descriptors (independent variable) impact the effectiveness of speech therapies (dependent variable). Baseline proficiency moderates this relationship, while the correlation between stutter descriptors and therapy effectiveness serves as a mediating factor. The goal is to discern the complex dynamics and provide personalized therapy recommendations for individuals with stuttering. | | | | | | | | |
| **Input and Output** | | | **Feature of This Solution** | | | | **Contribution & The Value of This Work** |
| |  |  | | --- | --- | | **Input** | **Output** | | Recorded speech audio | Quantitative analysis of the type and severity of stuttered disfluencies in speech and recommendations | | | | The various features in the solution are Stutter assessment using Mel-frequency Cepstral Coefficients (MFCC) features, therapy suggestion based on patient performance, separate Gated Recurrent CNN (GRCNN) models for detecting prolongation and repetition, an SVM model for recommending suitable therapies, and integration into a mobile application for personalized speech therapy. | | | | It based on Deep Learning uses Mel-frequency Cepstral Coefficients (MFCC) features for stutter assessment and recommends appropriate therapies based on patient performance. The system's value lies in automating tasks, improving accuracy, and providing an affordable solution for people who stutter. |
| **Positive Impact of this Solution in This Project Domain** | | | | | **Negative Impact of this Solution in This Project Domain** | | |
| By automating stutter assessment and recommending personalized therapies, the system offers an affordable and accessible solution. This technological intervention addresses challenges such as the high cost of private speech therapy and the need for customized treatment plans | | | | | the implementation of technology in stutter therapy raises concerns. Affordability and accessibility issues may persist, especially in regions with limited technological infrastructure. Privacy concerns also emerge as continuous monitoring and analysis of speech patterns become integral to therapy. | | |
| **Analyse This Work By Critical Thinking** | | | | **The Tools That Assessed this Work** | | | **What is the Structure of this Paper** |
| This work would be to include a larger and more diverse dataset of stuttered speech recordings to improve the accuracy and generalizability of the models. It provided a highly innovative and valuable in addressing the need for technology in speech therapy and providing an affordable and accessible solution for people who stutter. | | | | The tools used to assess this work include Python and its libraries (such as PyDub and Librosa) for audio processing and feature extraction, scikit-learn for training the SVM model, Keras with TensorFlow backend for developing and training the GRCNN models, and various deep learning models such as CNN and RNN. | | | Abstract   1. Introduction 2. Previous Work 3. Our Approach 4. Evaluation 5. Key Findings 6. Results 7. Conclusion 8. Future Scope |
| **Diagram/Flowchart** | | | | | | | |
|  | | | | | | | |

**--End of Paper 3--**

|  |
| --- |
| **4** |
| **Reference in APA format** | |  | | | | | |
| **URL of the Reference** | | **Authors Names and Emails** | | | | **Keywords in this Reference** | |
| <https://scholar.google.com/scholar?start=30&q=stuttering+recognition+machine+learning&hl=en&as_sdt=0,5#d=gs_qabs&t=1700053247762&u=%23p%3D6odTPCRxfVIJ> | | Seema Barda | | | | Stuttering, Speech dysfluency, Automatic Speech Recognition (ASR), Stuttering severity, Mel frequency Cepstral Coefficients (MFCC), Support Vector Machine (SVM) | |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc )** | | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | | | | **What are the components of it?** | |
| It uses Automatic Speech Recognition (ASR), Mel frequency Cepstral Coefficients (MFCC) for feature extraction, and Support Vector Machine (SVM) for feature classification in the context of analyzing stuttering severity. | | The solution aims to automate the assessment of stuttering severity by proposing an approach that uses Automatic Speech Recognition (ASR), focusing on features like Mel frequency Cepstral Coefficients (MFCC) and employing Support Vector Machine (SVM) for classification. | | | | The solution involves using Mel frequency Cepstral Coefficients (MFCC) for feature extraction and Support Vector Machine (SVM) for classification to automate the assessment of stuttering severity. | |
| **The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process** | | | | | | | |
| |  |  |  |  | | --- | --- | --- | --- | |  | **Process Steps** | **Advantage** | **Disadvantage (Limitation)** | | **1** | The system begins by pre-processing the input speech signal, including noise reduction and filtering. It then uses Voice Activity Detection to identify speech segments | The system enables real-time recognition of emotions in continuous speech, facilitating immediate and responsive interactions in human-computer interfaces. | Developing and implementing a real-time speech emotion recognition system, especially one based on advanced techniques like CNNs, may require specialized expertise and resources. | | **2** | Extract crucial speech features like MFCCs and spectral attributes, and employ a trained machine learning model, potentially a Convolutional Neural Network (CNN), for efficient emotion classification. | By understanding and classifying emotional states, the system can deliver personalized responses, enhancing user experience and engagement | The system may struggle with recognizing emotions not well-represented in the training set. | | **3** | A statistical analysis on recognized emotions, offering quantitative insights into the prevalence and distribution of different emotional states in the continuous speech data. |  |  | | | | | | | | |
| **Major Impact Factors in this Work** | | | | | | | |
| |  |  |  |  | | --- | --- | --- | --- | | **Dependent Variable** | **Independent Variable** | **Moderating variable** | **Mediating (Intervening ) variable** | | Stuttering Severity | Automated Speech Recognition System | Speech-language pathologists' expertise level | Efficiency of the Automated Speech Recognition System | | | | | | | | |
| |  | | --- | | **Relationship Among The Above 4 Variables in This article** | | The independent variable is the "Automated Speech Recognition System," which directly influences the dependent variable, the "Stuttering Severity Assessment." The efficiency of the recognition system acts as a mediating factor, impacting the accuracy of stuttering assessment. Concurrently, the expertise of "Speech-language pathologists" moderates this relationship, influencing how the system's effectiveness varies based on the pathologists' proficiency. | | | | | | | | |
| **Input and Output** | | | **Feature of This Solution** | | | | **Contribution & The Value of This Work** |
| |  |  | | --- | --- | | **Input** | **Output** | | Speech | Severity | | | | This solution utilizes Mel frequency cepstral coefficients (MFCC) for efficient feature extraction and employs Support Vector Machine (SVM) for classification. By analyzing spectral and temporal features, the system distinguishes between normal and stuttered speech, contributing to speech pathology by automating severity assessment. | | | | This work significantly contributes to the field of speech pathology by introducing an automated solution for assessing the severity of stuttering. By leveraging advanced techniques like MFCC and SVM, the system provides an efficient and accurate method, reducing the time-consuming manual efforts. The automation not only enhances the speed of assessment but also allows professionals to focus more on therapeutic strategies. |
| **Positive Impact of this Solution in This Project Domain** | | | | | **Negative Impact of this Solution in This Project Domain** | | |
| The automated stuttering severity assessment system offers significant positive impacts by streamlining diagnostic processes, reducing evaluation time, and enabling speech-language pathologists to concentrate on effective therapeutic interventions, ultimately enhancing patient outcomes and treatment quality. | | | | | There are challenges in terms of accuracy and reliability, as it heavily relies on speech recognition technology. The reliance on technology may limit the system's effectiveness in capturing the nuanced and subjective aspects of stuttering. | | |
| **Analyse This Work By Critical Thinking** | | | | **The Tools That Assessed this Work** | | | **What is the Structure of this Paper** |
| It should focus on addressing challenges and assessing the real-world implications of relying solely on automated systems for a nuanced condition like stuttering, ensuring it complements rather than replaces human expertise in therapeutic interventions | | | | Mel Frequency Cepstral Coefficients (MFCC) for feature extraction and Support Vector Machine (SVM) for classification. | | | * 1. Abstract   2. Introduction   3. Challenges   4. Types of Speech Recognition Techniques   5. Methodology   6. Conclusion |
| **Diagram/Flowchart** | | | | | | | |
|  | | | | | | | |

**--End of Paper 4—**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Version 2.0 Week 2** | | | | | | | |
| **5** |
| **Reference in APA format** | |  | | | | | | |
| **URL of the Reference** | | **Authors Names and Emails** | | | | **Keywords in this Reference** | | |
| <https://www.sciencedirect.com/science/article/pii/S2665917423002490> | | M. Mahendran  R. Visalakshi  S. Balaji | | | | Dysarthia, Speech detection, CNN, MFCC Feature extraction | | |
| **The Name of the Current Solution (Technique/ Method/ Scheme/ Algorithm/ Model/ Tool/ Framework/ ... etc )** | | **The Goal (Objective) of this Solution & What is the problem that need to be solved** | | | | **What are the components of it?** | | |
| The paper proposes a CNN-based model that uses various speech features to detect dysarthria in patients | | The goal of the proposed solution is to detect dysarthria in patients using a CNN-based model that analyses various speech features. The problem that needs to be solved is the difficulty in diagnosing dysarthria, which is speech impairment caused by various underlying conditions. | | | | A Convolutional Neural Network-based model to detect dysarthria in patients. The model analyses various speech features such as zero crossing rates, MFCCs, spectral centroids, and spectral roll off. The TORGO speech signal database is used for training and testing the model. | | |
| **The Process (Mechanism) of this Work; Means How the Problem has Solved & Advantage & Disadvantage of Each Step in This Process** | | | | | | | | |
| |  |  |  |  | | --- | --- | --- | --- | |  | **Process Steps** | **Advantage** | **Disadvantage (Limitation)** | | **1** | Pre-processing the data, this includes silence removal and noise reduction. | The proposed solution is generalizable across languages, making it useful for detecting dysarthria in patients speaking different languages. | It requires a large amount of data for training the model, which may not always be available. | | **2** | Feature extraction using various speech features such as zero crossing rates, MFCCs, spectral centroids, and spectral roll off. | The end-to-end framework optimizes feature extraction, distance matrix computation, and classification, making the model efficient and effective | The model's accuracy may be affected by variations in speech patterns due to factors such as age, gender, and accent. | | **3** | Training the CNN model using the TORGO speech signal database. | It surpasses state-of-the-art CNN-based systems, according to experimental results on two dysarthric speech datasets | The solution may not be suitable for detecting dysarthria in patients with severe speech impairments, as the model may not be able to capture the unique characteristics of their speech. | | **4** | Testing the model on the same database to evaluate its accuracy in detecting dysarthria. |  |  | | **5** | Using the trained model to detect dysarthria in patients by analysing their speech features. |  |  | | | | | | | | | |
| **Major Impact Factors in this Work** | | | | | | | | |
| |  |  |  |  | | --- | --- | --- | --- | | **Dependent Variable** | **Independent Variable** | **Moderating variable** | **Mediating (Intervening ) variable** | | Dysarthria Diagnosis | Severity level of dysarthria | Age | Speech features | | | | | | | | | |
| |  | | --- | | **Relationship Among The Above 4 Variables in This article** | | This study examines the link between dysarthria severity and the accuracy of diagnosis using a Convolutional Neural Network . Speech features like zero crossing rates and MFCCs serve as mediating variables, explaining how dysarthria severity affects diagnosis accuracy. Age is considered as a potential moderating variable, suggesting its influence on this relationship. Overall, the study showcases the CNN's effectiveness in early dysarthria diagnosis, considering severity levels and potential moderating factors. | | | | | | | | | |
| **Input and Output** | | | **Feature of This Solution** | | | | **Contribution & The Value of This Work** | |
| |  |  | | --- | --- | | **Input** | **Output** | | An audio sample of speech | A binary classification label, indicating normal speech or dysarthria speech | | | | The proposed solution uses a Convolutional Neural Network-based model for dysarthria detection, which has an accuracy score of 93.87%. The model uses various speech features such as zero crossing rates, MFCCs, spectral centroids, and spectral roll off for the analysis of speech signals | | | | The contribution of this work is the development of a CNN-based model for dysarthria detection, which has a high accuracy score and uses various speech features for analysis. The value of this work lies in its potential to improve the quality of life for individuals with dysarthria by enabling early detection and intervention. | |
| **Positive Impact of this Solution in This Project Domain** | | | | | **Negative Impact of this Solution in This Project Domain** | | | |
| It can aid in early detection and better management of dysarthria, leading to improved quality of life for individuals with the impairment. Additionally, the proposed model shows promising results in detecting dysarthria with high accuracy, which can potentially reduce the need for invasive diagnostic procedures. | | | | | It may not be accessible to individuals who do not have access to the necessary technology or resources. Additionally, the model's effectiveness may be limited to the specific dysarthric speech datasets used in the study, and further research is needed to validate its effectiveness on a larger scale. | | | |
| **Analyse This Work By Critical Thinking** | | | | **The Tools That Assessed this Work** | | | **What is the Structure of this Paper** | |
| This work is a significant step towards improving the quality of life for individuals with dysarthria, but further research is needed to validate the proposed model's effectiveness. | | | | The tools used to assess this work include a pairwise distance-based CNN, feature extraction, distance matrix computation, and classification. The model was evaluated on two dysarthria speech datasets, and the results showed that it outperformed state-of-the-art CNN-based systems. | | | 1. Introduction  2. Related Work  3. Materials and Methods  3.1. Data Collection  3.2. Data Preprocessing  3.3. Feature Extraction  3.4. Convolutional Neural Network  3.5. Model Training and Evaluation  4. Results and Discussion  5. Conclusion and Future Work  6. References | |
| **Diagram/Flowchart** | | | | | | | | |
|  | | | | | | | | |

**Work Evaluation Table**

**<Use the same factors you have used in "Work Evaluation Table" to build your own "Proposed and Previous comparison table ">**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Work Goal** | **System's Components** | **System's Mechanism** | **Features /Characteristics** | **Cost** | **Speed** | **Security** | **Performance** | **Advantages** | **Limitations /Disadvantages** | **Platform** | **Results** |
| **Hadhami Aouani**  **Yassine Ben Ayed** **(2020)** | To create speech emotion recognition system | An emotion recognition system based on speech signals in two-stage approach, namely feature extraction and classification engine. | The system involves a two-stage approach, first stage focusing on feature extraction using a combination of MFCC, ZCR, TEO, and HNR. The second involves dimension reduction using an auto-encoder and classification using a support vector machine (SVM). | It utilizes 39 Mel Frequency Cepstral Coefficients (MFCC), Zero Crossing Rate (ZCR), Harmonic to Noise Rate (HNR), and Teager Energy Operator (TEO) as audio features for emotion recognition. Additionally, an auto-encoder is employed for dimension reduction to select pertinent parameters from the extracted features. |  |  |  | The model achieved an accuracy rate of 74.07% for recognizing six emotions from the RML dataset. The results showed that the RBF kernel of SVM outperformed the linear and polynomial kernels in terms of recognition rates. | * **1.** Improved identification rates compared to other systems. * **2.** Utilization of auto-encoder for feature selection. | **1.** Limited dataset size  **2.** Empirical tuning of SVM parameters |  | The solution achieved a maximum identification rate of 74.07% for recognizing six emotions using the proposed feature extraction, dimension reduction with auto-encoder, and SVM classification. |
| **Husbaan I. Attar, Nilesh K. Kadole, Omkar G. Karanjekar, Devang R. Nagarkar, Prof. Sujeet (2022)** | The main objective is to improve a person's speech fluency by accurately diagnosing stutter and suggesting appropriate training exercises for practice. | It consists of Voice Activity Detection to identify speech segments, Speech Segmentation for meaningful division, Signal Pre-Processing for conditioning audio, Feature Extraction extracting relevant speech features, Emotion Classification utilizing machine learning, and Statistics Analysis of Emotion Frequency for insight. | The proposed speech emotion recognition system uses machine learning techniques to classify emotions expressed in continuous speech. The system includes feature extraction, feature selection, and classification stages to achieve high accuracy rates in real-time recording experiments. | The important features used in the proposed system include Log-Mel Spectrogram, Mel-Frequency Cepstral Coefficients (MFCCs), pitch, and energy. |  |  |  | The proposed system achieved high accuracy rates in emotion classification using methods such as Long Short Term Memory (LSTM), Convolutional Neural Networks (CNNs), Hidden Markov Models (HMMs), and Deep Neural Networks (DNNs). | **1.** High accuracy rates achieved in real-time recording experiments.  **2.** It has potential applications in various fields. | **1.** Limited to the four emotions considered in the study.  **2.** It may not generalize well to other languages or cultures. |  | The final result of the solution demonstrated effective speech emotion recognition with potential applications in fields such as mental state assessment in dangerous environments and customer satisfaction monitoring. |
| **Dr. Mrs. Gresha Bhatia, Binoy Saha, Mansi Khamkar, Ashish Chandwani , Reshma Khot (2021)** | The main objective is to improve a person's speech fluency by accurately diagnosing stutter and suggesting appropriate training exercises for practice. | The components are stutter assessment, therapy suggestion, Gated Recurrent CNN (GRCNN) models, SVM model, and a mobile application. | The system uses deep learning models, Gated Recurrent CNN and SVM, to automate stutter assessment and personalize therapy recommendations. | Its main features include speech fluency improvement, personalized treatment, and continuous performance monitoring. |  |  |  | The solution achieves high validation accuracies of approximately 95% for identifying prolongation and 92% for repetition in speech audio. | **1.** Improved speech fluency.  **2.** It provides personalized treatment. | **1.** Potential overreliance on technology.  **2.** There is a huge requirement for robust data. |  | The results demonstrate the system's effectiveness in accurately diagnosing stutter types and recommending appropriate therapies based on stutter descriptors and speech fluency improvement. |
| **Seema Badra (2019)** | The solution aims to automate the assessment of stuttering severity by proposing an approach that uses Automatic Speech Recognition (ASR), focusing on features like Mel frequency Cepstral Coefficients (MFCC) and employing Support Vector Machine (SVM) for classification. | The solution involves using Mel frequency Cepstral Coefficients (MFCC) for feature extraction and Support Vector Machine (SVM) for classification to automate the assessment of stuttering severity. | The solution involves an automatic recognition system that utilizes speech processing techniques to assess the severity of stuttering, aiming to reduce the manual workload of speech-language pathologists. | It incorporates feature extraction algorithms to analyze spectral and temporal features of speech, enabling the identification of repetitions and prolongations characteristic of stuttered speech. |  |  |  | The solution demonstrates promising performance in accurately recognizing and quantifying the rate of stuttering severity, showcasing its potential to enhance the efficiency of assessment processes. | **1.** It provides an automation of stuttering severity assessment there by reducing manual workload.  **2.** It provides comprehensive evaluation of speech patterns leading to more accurate assessments. | **1.** The need for rigorous validation and calibration for accuracy.  **2.** It has potential limitations in capturing nuanced aspects of stuttering. |  | The final results indicate a significant reduction in the time required for assessing stuttering severity, thereby improving the overall workflow of speech-language pathologists and potentially leading to more timely interventions for individuals with speech disorders. |
| **M. Mahendran, R. Visalakshi, S. Balaji (2021)** | The goal of the proposed solution is to detect dysarthria in patients using a CNN-based model that analyses various speech features. The problem that needs to be solved is the difficulty in diagnosing dysarthria, which is speech impairment caused by various underlying conditions. | A Convolutional Neural Network-based model to detect dysarthria in patients. The model analyses various speech features such as zero crossing rates, MFCCs, spectral centroids, and spectral roll off. The TORGO speech signal database is used for training and testing the model. | The proposed solution utilizes a pairwise distance-based CNN to compare frame-level distance patterns between healthy and dysarthric speech representations, achieving high accuracy in dysarthric speech identification . | The main features of the solution include the use of phonetically-balanced AP representations, an end-to-end framework for feature extraction and classification, and generalizability across languages, surpassing state-of-the-art CNN-based systems in dysarthric speech identification |  |  |  | The proposed CNN-based model achieves an accuracy score of 93.87% in early dysarthric speech diagnosis, demonstrating promising results in detecting dysarthria | **1.** Generalizable across languages for dysarthria detection.  **2.** It provides end-to-end framework with optimized feature extraction, distance matrix computation, and classification. | **1.** Overfitting is a major concern.  **2.** Valid padding reduces the number of features. |  | The final results show that the solution outperforms traditional CNN models and achieves high accuracy in dysarthric speech identification, indicating its potential for effective impairment management. |

**© Dr Abeer Alsadoon 2016\_CSU Sydney Study Centre**