PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
Ministry of Higher Education and Scientific Research

UNIVERSITY OF ECHAHID HAMMA LAKHDAR-EL OUED

FACULTY OF EXACT SCIENCES
Computer Science Department

Thesis
submitted in partial fulfillment of the requirements for the degree of

ACADEMIC MASTER

Field Mathematics and Computer Science
Option Computer Science
Specialty Distributed Systems and Artificial Intelligence

Presented by :

- Nariman Chetioui
- Zaouche Fatma

Title

THE USE OF ARTIFICIAL INTELLIGENCE TO DETECT SPEECH CHARACTERISTICS IN ELEMENTARY SCHOOL CHILDREN WITH SPEECH DISORDERS

Supported on xx-xx-2024 before the jury:

Examination Commitee:

Dr. Mohamed Lamine MEKHALFI	MAA	President
Dr. Mohammd Anouar NAOUI	MAA	Examiner
Dr. Adel BERHOUM	MAA	Supervisor
Prf. Mohammad Charaf Eddine MEFTAH	MCA	Assistant Supervisor

Academic year: 2023/2024

Dedication

After years of hard work, I now stand on the brink of my graduation, proud and grateful. I thank Allah for His guidance and support throughout this journey. To the pride of our lives, to that pair for whom words and phrases are not enough to express my praise and gratitude: to you, my mother, and to you, my father.
You have endured as much, if not more, struggle as we have during this academic journey, and you have been unwavering pillars of support and our greatest advocates. We extend our deepest thanks and gratitude to you first. To my siblings, for their invaluable guidance.
I also dedicate this moment to my friends who stood by me during tough times. Finally, I acknowledge that none of this would have been possible without Allah's blessings, which have brought immense joy and achievement to this significant day.

THANKS

In the beginning, we praise God Almighty for His guidance and grace, and we bear witness that there is no god but Allah, and He is alone without partner.

In all of this, and before anything else, we remember our Gaza. Thank you, Gaza, for teaching us the values of determination and possibility, and that miracles happen when there is faith, will, and determination. You are, by God, the miracle of this time!

At the end of this paragraph, we thank our wonderful parents for their faith in us and their continuous support. We carry with us the lessons of perseverance, resilience, and self-belief, and we promise to use them to make a positive impact in the world.

In gratitude and appreciation, we extend our sincere thanks to Professor

Mr.Berhoum Adel

for his valuable guidance and encouragement, and to all our esteemed professors in the Department of Computer Science at the University of Echahid Hamma Lakhdar, and to everyone who helped us make this work a success.

Résumé

Les troubles de la parole sont une condition courante de l'enfance qui peut affecter la capacité d'un enfant à communiquer efficacement. Les méthodes traditionnelles de diagnostic des troubles de la parole nécessitent une évaluation clinique par un orthophoniste (orthophoniste). Cependant, ces évaluations peuvent être longues et coûteuses, et elles ne sont pas toujours accessibles à tous les enfants.L'intelligence artificielle (IA) a le potentiel de révolutionner la façon dont les troubles de la parole sont diagnostiqués. Les systèmes basés sur l'IA peuvent être utilisés pour analyser automatiquement les enregistrements vocaux et identifier des schémas indicateurs d'un trouble de la parole. Cela pourrait conduire à une détection et un traitement plus précoces des troubles de la parole, ce qui pourrait améliorer les résultats à long terme des enfants.Cet article examine l'état actuel de la recherche sur l'utilisation de l'IA pour détecter les caractéristiques de la parole chez les enfants d'école primaire souffrant de troubles de la parole. L'article examine les différents types de techniques d'IA qui ont été utilisées à cette fin, ainsi que les défis et les limites de cette recherche.

Mots Clés: Speech disorder,SLP , SSD ,Early intervention, Diagnosis, Children

Abstract

Speech disorders are a common childhood condition that can affect a child's ability to communicate effectively. Traditional methods of diagnosing speech disorders require a clinical assessment by a speech-language pathologist (SLP). However, these assessments can be time-consuming and expensive, and they may not always be accessible to all children.Artificial intelligence (AI) has the potential to revolutionize the way speech disorders are diagnosed. AI-powered systems can be used to automatically analyze speech recordings and identify patterns that are indicative of a speech disorder. This could lead to earlier detection and treatment of speech disorders, which could improve children's long-term outcomes.This paper reviews the current state of research on the use of AI to detect speech characteristics in elementary school children with speech disorders. The paper discusses the different types of AI techniques that have been used for this purpose, as well as the challenges and limitations of this research.

Keywords: Speech disorder,SLP , SSD ,Early intervention, Diagnosis, Children .

ملخص

اضطرابات الكلام هي حالة شائعة في مرحلة الطفولة يمكن أن تؤثر على قدرة الطفل على التواصل بشكل فعال. تتطلب الطرق التقليدية لتشخيص اضطرابات الكلام تقييماً سريرياً من قبل أخصائي أمراض النطق واللغة (SLP). ومع ذلك ، يمكن أن تكون هذه التقييمات تستغرق وقتاً طويلاً ومكلفة ، وقد لا تكون دائماً متاحة لجميع الأطفال. يمتلك الذكاء الاصطناعي (AI) القدرة على إحداث ثورة في طريقة تشخيص اضطرابات الكلام. يمكن استخدام الأنظمة التي تعمل بالذكاء الاصطناعي لتحليل تسجيلات الكلام تلقائياً وتحديد الأنماط التي تشير إلى اضطراب الكلام. يمكن أن يؤدي هذا إلى الكشف المبكر وعلاج اضطرابات الكلام ، مما قد يحسن نتائج الأطفال على المدى الطويل. تستعرض هذه الورقة البحثية الحالة الراهنة للبحث حول استخدام الذكاء الاصطناعي للكشف عن خصائص الكلام لدى طلاب المدارس الابتدائية الذين يعانون من اضطرابات الكلام. تناقش الورقة أنواع تقنيات الذكاء الاصطناعي المختلفة التي تم استخدامها لهذا الغرض ، بالإضافة إلى التحديات والقيود التي تواجه هذا البحث.

كلمات مفتاحية: اضطراب النطق، أخصائي أمراض النطق واللغة SLP ، SSD ،التدخل المبكر، التشخيص، الأطفال.

Table des matières

Liste des figures

Liste des tableaux

General Introduction

Speech issues are quite common, in kids affecting around 8% of school children. These problems come in forms. Can really make it tough for a child to communicate effectively. Normally identifying speech issues relies on the expertise of Speech Language Pathologists (SLPs) who use methods like observation, listening and tests. Unfortunately this process can be time consuming and demanding.
Artificial intelligence (AI) is showing promise in changing how we diagnose speech problems. AI tools can analyze speech recordings automatically potentially spotting signs of speech issues quickly[37].

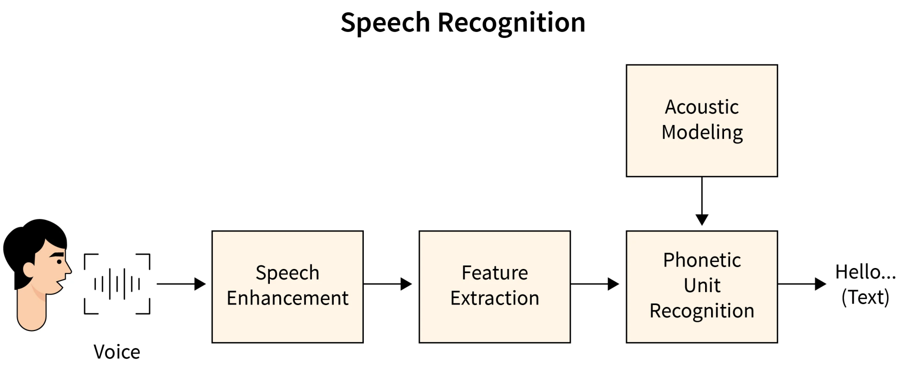

Figure 0.1: Speech Recognition in AI.

Different AI techniques can be used for this purpose with machine learning being a choice. By training machine learning models on datasets that include speech recordings from children with and without speech problems these models can learn to identify patterns linked to speech disorders. Once trained these models can determine if new speech recordings are normal or show signs of disorder.
This thesis is structured into four comprehensive chapters, with each chapter dedicated to exploring a distinct artificial intelligence technique designed for the identification of speech traits associated with various speech disorders.

- The first chapter provides an introduction to the field of speech disorder detection, outlining the significance of early and accurate diagnosis and the role artificial intelligence can play in this domain. It sets the stage for the subsequent chapters by presenting the thesis objectives and the rationale behind choosing artificial intelligence methods for this research.
- Chapter two delves into the first artificial intelligence method, detailing the algorithm's design, implementation, and the specific speech characteristics it targets. This chapter also includes a review of the existing literature, highlighting the advancements in the field and how this method builds upon or differs from previous work.
- The third chapter continues with the second artificial intelligence method, focusing on a different set of speech characteristics. This chapter presents the theoretical framework, the data collection process, and the method's effectiveness in detecting speech disorders through various tests and experiments.
- Finally, the fourth chapter discusses the last two artificial intelligence methods, comparing and contrasting their approaches to speech disorder detection. It includes a comprehensive analysis of the results obtained from applying these methods, discussing their strengths, limitations, and potential for real-world application.
- In conclusion, the thesis synthesizes the findings from all four chapters, offering insights into the overall performance of the artificial intelligence methods studied and suggesting directions for future research in the field of speech disorder detection using AI technologies.

0.1 General Work Context:

Speech disorders in children represent a significant challenge, affecting an estimated 8% of school-age children [25] [1]. These disorders encompass a wide range of conditions, including articulation disorders, phonological disorders, fluency disorders (such as stuttering), and voice disorders. Traditionally, the diagnosis of speech disorders has relied on the expertise of speech-language pathologists (SLPs) [29]. SLPs employ a variety of methods to assess speech disorders, including:

- **Observation:** Watching and analyzing a child's speech patterns during conversation or structured tasks.
- **Increased Accessibility:** Listening: Listening to the quality of a child's speech, including pitch, resonance, and fluency.
- **Testing:** Administering standardized tests or assessments to evaluate specific aspects of speech production, such as articulation or phonological skills.

While these traditional diagnostic methods are valuable, they can be time-consuming and labor-intensive. Furthermore, the subjective nature of human observation and interpretation may introduce variability into the diagnostic process.

0.2 The thesis objectives:

The primary aim of this thesis is to investigate and evaluate the application of artificial intelligence (AI) techniques in diagnosing speech disorders among children. Specifically, the thesis aims to explore how AI-powered tools can enhance and streamline the diagnostic process, potentially improving accuracy, efficiency, and accessibility.
By leveraging AI techniques, such as machine learning algorithms, deep neural networks, and convolutional neural networks, the thesis seeks to develop automated systems capable of analyzing speech recordings and identifying potential signs of speech disorders. This approach has the potential to supplement traditional diagnostic methods used by SLPs, offering a more objective and data-driven approach to diagnosis.

0.3 Contributions:

This thesis offers several potential contributions to the field of speech pathology, particularly in streamlining the diagnostic process for speech disorders in children:

- **Automation of Diagnosis:** By exploring the application of AI techniques, such as machine learning algorithms and neural networks, the thesis aims to automate aspects of the diagnostic process for speech disorders. This automation has the potential to reduce the time and effort required for diagnosis, allowing for more efficient evaluation and treatment planning.
- **Objective Assessmen:** AI-powered diagnostic tools have the potential to provide more objective assessments of speech disorders compared to traditional methods that rely heavily on subjective interpretation by clinicians. By analyzing speech recordings and identifying patterns associated with speech disorders, these tools can offer more standardized and consistent assessments.
- **Enhanced Accuracy:** The use of AI techniques, particularly deep learning algorithms, may lead to improved accuracy in diagnosing speech disorders. These algorithms have the ability to learn complex patterns and relationships within speech data, potentially leading to more accurate and reliable diagnoses compared to traditional methods.
- **Increased Accessibility:** By developing automated AI-powered diagnostic tools, the thesis aims to increase accessibility to speech disorder diagnosis, particularly in underserved communities or regions with limited access to trained speech-language

pathologists. These tools have the potential to be deployed in a variety of settings, including schools, clinics, and remote healthcare facilities.

- **Facilitation of Early Intervention:** Early detection and intervention are crucial for improving outcomes for children with speech disorders. By streamlining the diagnostic process and providing timely assessments, AI-powered tools developed in this thesis may facilitate earlier intervention and support for affected children, leading to better long-term outcomes.

CHAPTER 1:

State of the art

Chapitre 1

State of the art

Introduction

Speech disorders are common among children in primary school and younger as they can affect a child's ability to speak clearly, communicate effectively, and understand speech. According to research and statistics, about 5% of first-grade children have noticeable speech disorders, including stuttering, articulation disorders, and phonological disorders. Speech disorders are the most common type of communication disorder in this age group. The types of speech disorders in elementary school students include articulation disorders, phonological disorders, and stuttering. Articulation disorders are the most common, followed by phonological disorders and stuttering. Most speech disorders in elementary school students do not have a known cause, but there may be genetic or environmental factors that play a role in their development.

Early detection of speech disorders in elementary school children is crucial for improving communication skills, academic success, social interaction, emotional well-being, language development, and identifying underlying issues. However, assessing speech and language disorders can present various challenges for speech-language pathologists (SLPs), such as the lack of clear diagnostic guidelines, developmental differences, cultural and linguistic considerations, limited assessment tools, and the dynamic nature of communication disorders.

1.1 Impact of speech disorders on physical-acoustic speech

Speech disorders have a profound impact on the physical-acoustic aspects of speech production, particularly among children in primary school. Understanding the context of these disorders is crucial in recognizing their effects and importance for early detection.

1.1.1 Define the context of the speech disorders

The context of speech disorders refers to the specific conditions and characteristics that surround and define these disorders.

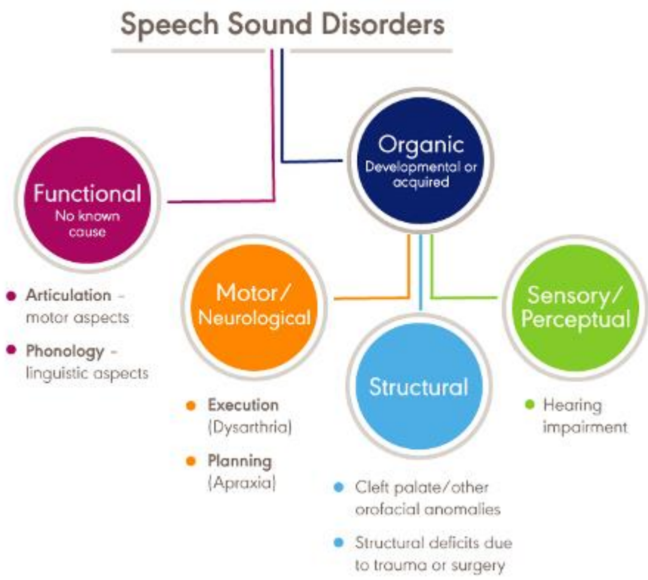

Figure 1.1: Speech Disorders Medical Classification (ASHA).[**?**]

It includes various aspects such as the types of speech disorders, their symptoms, causes, and treatment options. Here is a breakdown of the context of speech disorders :

1.1.1.1 Definition of speech disorders

"A disorder that manifests as the individual's lack of ability to communicate verbally with others in a manner that is consistent with their chronological and mental age. This results from the substitution of some letters for others, the omission of some letters, or the distortion of the sound of speech, which leads to the inability to pronounce speech properly and, consequently, hinders the understanding of others for this speech." [28]
Procedurally, it means the inability of Primary School students to pronounce some speech sounds, which appears in one or more of the following disorders : substitution of the pronunciation of one sound instead of another, deletion of one or more sounds from the word, or distortion (distortion of the speech sound in a way that brings it closer to the original sound but is not exactly the same, or addition of one or more sounds

to the spoken word, these manifestations are speech sound disorders or pronunciation substitution deletion - distortion, distortion or heart - addition).

1.1.1.2 Speech disorders and physical-acoustic speech

Speech disorders manifest in various physical-acoustic speech characteristics that can be analyzed to understand and diagnose these conditions. In speech disorders, issues with pitch, intensity, duration, and spectral features can be observed:

- **Pitch:** Speech disorders can affect the pitch of speech, leading to abnormalities in the fundamental frequency (F0) of the voice. For instance, stuttering can result in dysfluencies that impact the pitch of speech [26] [46]
- **Intensity:** Changes in intensity, which refers to the loudness of speech, can be indicative of speech disorders. For example, individuals with positive symptoms of schizophrenia-spectrum disorders may exhibit differences in vocal intensity compared to those with negative symptoms [17]
- **Duration:** Speech disorders can influence the duration of speech sounds, causing variations in the length of vowels and consonants. Dysfluent speech, as seen in stuttering, can result in shorter vowel durations compared to fluent speech [17]
- **Spectral Features:** Spectral features of speech, such as formant frequencies and bandwidth, can be altered in individuals with speech disorders. Differences in spectral parameters like F1 and F2 bandwidths have been observed in patients with positive symptoms of schizophrenia-spectrum disorders [26][17]

1.2 Types of disorders

1.2.1 Substitution:

Substitution errors in speech, as you mentioned, occur when a person produces an incorrect sound in place of the intended sound. This can be observed in children who may replace certain letters or sounds with others that are not appropriate in the context. For instance, a child might substitute the sound "sh" for the sound "s" or replace "t" with "w." [12].

1.2.2 Addition:

Addition occurs when a child adds an extra sound or syllable to a word, which can make their speech unclear and difficult to understand. These additional sounds or syllables are not part of the intended word and can disrupt the normal flow of speech.

For example, a child may say "sismaka" instead of "smack" or "mamrouha" instead of "mama." These additions can make it challenging for others to comprehend the child's speech, especially when they persist over time.
Repetition of syllables is another type of addition error. A child may repeat one or more syllables within a word or across multiple words. For instance, they might say "wawa" instead of "water" or "dada" instead of "dog." This repetition of syllables can also impact the clarity and intelligibility of the child's speech [12].

1.2.3 Distortion:

is another type of speech error that occurs when a child pronounces sounds incorrectly. In this case, the child attempts to produce a sound but is unable to do so accurately because they do not use the correct method of exhaling the airflow needed to produce the sound. This results in a distorted or altered pronunciation of the word.
For example, a child may pronounce the word "madrasa" as "mudarathah" due to the distortion of certain sounds. Distortion errors can occur in various sounds or phonemes and can affect both children[12].

1.3 Possible causes

- **Genetic Factors:** Some speech and language disorders have a genetic basis, meaning they can run in families. Genetic conditions such as Down syndrome, Fragile X syndrome, and specific language impairment (SLI) are known to affect speech and language development in children [39] [27].
- **Neurological Factors:** Certain neurological conditions, such as cerebral palsy or brain injuries, can affect the areas of the brain responsible for speech and language production. These conditions can result in difficulties with articulation, language comprehension, and expressive language skills[21].
- **Hearing Loss:** Hearing plays a crucial role in speech and language development. Children with hearing loss may have difficulty perceiving and producing speech sounds, which can impact their language skills. Hearing loss can be caused by factors such as ear infections, genetic conditions, or exposure to loud noises[19].
- **Environmental Factors:** Environmental factors can also contribute to speech and language disorders in children. Exposure to toxins, such as lead or mercury, during pregnancy or early childhood can have detrimental effects on speech and language development. Additionally, a lack of exposure to language-rich environments or limited opportunities for communication can hinder language acquisition [27].
- **Psychosocial Factors:** Emotional stress, neglect, abuse, or inconsistent language input can negatively impact a child's speech and language development[45].

- **Developmental Delays:**Children who experience delays in their overall development may also have delays in speech and language skills. These delays can be caused by a variety of factors such as prematurity, low birth weight due to underdeveloped organs, including the brain, or other developmental disorders[19] [56].

1.4 Traditional diagnosis of speech disorders

This process (diagnosing speech and language disorders) goes through several stages

1.4.1 Case history:

Case history, or medical history, refers to the collection of personal and historical information about an individual from birth to the present time. It includes information about pregnancy, maternal health, and the individual's development in various aspects such as motor skills, cognitive abilities, and language development. It also encompasses details about the individual's medical history and family history [52]. Case history is an important component of assessing an individual's health and well-being. It helps in understanding genetic and environmental factors that may impact their health. Information gathered in a case history is typically obtained from parents.

1.4.2 Child observation:

Child observation is done by monitoring and noting the child's interactions and behaviors during their interactions with family members, siblings, peers, or in the school setting. It can be carried out by teachers, classmates, or through activities within the classroom, such as reading, oral expression, and note-taking [40].
Through child observation, information is gathered about the child's social interactions, language and communication skills, motor skills, attention and concentration levels, and their response to teaching and guidance. This helps in assessing the child's development and identifying any difficulties or special needs they may have.

1.4.3 Assessment of communication skills

The assessment of communication skills is typically conducted by speech and language specialists. It involves gathering spontaneous speech samples from the learner and administering appropriate standardized language tests based on their age and developmental stage. Samples of the learner's speech are collected both at home and in school settings. Once the data is collected, the specialist studies and analyzes the language samples to determine the learner's strengths and weaknesses[**?**].

1.4.4 Referral to Other Specialists

Many specialists work within a multidisciplinary team framework. When a specialist does not work in a team, they are the primary person responsible for their work [59]. If they suspect that a learner has a hearing problem or a developmental delay, they will refer them to the appropriate specialist. Diagnostic tools used for speech disorders in children can be divided into four categories:

1.4.4.1 Basic Tools

Clinical examination of the speech organs, movement, and hearing to identify any structural or functional problems that may affect speech,and acoustic analysis to assess the quality of the child's voice and communication. [9].

1.4.4.2 Assistive Tools

Use of the latest technology and devices to record and analyze speech and communication.[7].

1.4.4.3 Functional Assessment

Involves a set of standardized tests to measure various aspects of a child's learning and speech abilities.[7]
Examples of functional assessment tools include:

- **Templin-Darley Test:**This test assesses the child's articulation skills.
- **Triota Ten Word Test:**This test assesses the child's speech sound production.

1.4.4.4 Advanced Tools:

Standardized tools used to study the cause of a problem in depth and understand it better.[7].

1.4.5 Impact speech disorders for primary school children

Speech disorders can significantly impact primary school children in various aspects of their academic, social, and emotional development. Here are some potential impacts:

1. **Academic Challenges:**
    - **Reading, Writing, and Spelling:**Children with speech and language disorders may struggle with reading, writing, and spelling. Their ability to comprehend and process verbal information can be compromised, affecting their completion of tasks and assignments as expected.[47]

- **Educational Achievement:**These challenges can hinder educational achievement, potentially impacting their future lives and society.[9]

2. **Social Interaction:**
   - **Self-Esteem and Confidence:** Difficulties communicating within the classroom may affect a student's self-esteem and self-confidence.
   - **Interpersonal Relationships:** Language skills are crucial for developing and maintaining social relationships. Children with speech and language disorders may find it challenging to interact with peers and form connections.[9]

3. **Support and Frustrations:**
   - **Teachers' Role:**Teachers play a significant role in supporting children's development. However, they may be unprepared for addressing the needs of children with speech sound disorders.
   - **Limited Support:**Both parents and teachers report limited support for children with speech sound disorders in schools and classrooms.[9]

1.4.6 Importance of early detection

Early detection of speech disorders in elementary school children is crucial for several reasons:

- **Improved Communication Skills:** Early identification of speech disorders allows for timely intervention and treatment, increasing the chances of improving a child's communication skills [8]. Speech and language learning are essential for academic success, social interactions, and overall well-being.

- **Academic Success:** Language skills are closely linked to academic success. Children with speech and language disorders may struggle with reading, writing, and understanding instructions in the classroom[11].Early detection enables educators and speech-language pathologists (SLPs) to provide appropriate support and interventions to help children overcome speech difficulties and succeed in school [10].

- **Social Interaction:** Speech disorders can affect a child's ability to communicate effectively with peers and adults, leading to social isolation and difficulties in forming relationships. Early detection and intervention can help children develop the necessary communication skills to interact confidently and participate fully in social settings [10].

- **Emotional Well-being:** Children with speech disorders may experience frustration, low self-esteem, and anxiety due to their difficulties in expressing themselves. Early detection allows for early intervention, which can address these emotional challenges and provide support to improve a child's overall well-being [20].

- **Language Development:** Speech disorders can impact a child's language development, including understanding and using words, following directions, and expressing thoughts and ideas[10]. Early detection allows for early intervention, which can help children develop age-appropriate language skills and prevent further delays or difficulties [11].
- **Identifying Underlying Issues:** Early detection of speech disorders can also help identify any underlying issues that may be contributing to the speech difficulties. This may include hearing loss, developmental delays, or other related conditions. Identifying these underlying issues early on allows for appropriate referrals and comprehensive treatment [10].

1.5 Challenges in SLP Assessment

Assessing speech and language disorders in children in primary school and younger can present several challenges for speech-language pathologists (SLPs). These challenges can impact the accuracy and effectiveness of the assessment process.Here are some of the challenges faced in SLP assessment:

1.5.1 Lack of Clear Diagnostic Guidelines

The absence of clear diagnostic guidelines for identifying children with speech sound disorders (SSD) can lead to confusion . Differentiating between articulation disorders and phonological disorders can be challenging, as they involve different aspects of speech errors. Articulation disorders involve difficulties with the physical movements required for speech sound production, while phonological disorders involve difficulties with the underlying sound patterns of language [19].

1.5.2 Developmental Differences

Assessing speech and language disorders in individuals of different age groups, such as adults, children, and infants/toddlers, requires different approaches. Each age group has unique developmental milestones and communication needs, which must be taken into account when selecting assessment tools and interpreting results.

1.5.3 Cultural and Linguistic Considerations

Cultural and linguistic factors can significantly impact the assessment process. Standardized assessments may not always be linguistically and culturally appropriate for individuals from diverse backgrounds. Translating standardized assessments can invalidate the results [21]. SLPs need to consider the cultural and linguistic backgrounds of their clients to ensure accurate assessment and diagnosis.

1.5.4 Limited Assessment Tools

While there are various assessment tools available for SLPs, there may be limitations in terms of their applicability and effectiveness for specific populations or disorders. SLPs need to carefully select assessment tools that are appropriate for the individual's age, language profile, cognitive functioning, and hearing impairment [21].

1.5.5 Dynamic Nature of Communication Disorders

Communication disorders can be complex and dynamic, with symptoms and abilities changing over time. This dynamic nature can make it challenging to accurately assess and diagnose individuals with speech and language disorders. SLPs need to consider the individual's progress and monitor their response to intervention over time.

Conclusion

peech disorders can have a significant impact on primary school children, affecting their communication skills, academic performance, social interactions, and overall well-being. Children with speech disorders may face challenges in articulating sounds, expressing thoughts clearly, understanding instructions, and developing reading and writing skills. They may also experience social and emotional difficulties, such as low self-esteem and isolation. However, with early identification and appropriate intervention, many children with speech disorders can make significant progress and improve their communication skills.

CHAPTER 2:

Related work

Chapitre 2

Related work

Introduction

Speech disorders require early detection and diagnosis to create effective treatments and improve communication. Professionals use various tools including technology, learning models, and physical-acoustic analysis. Machine learning models analyze speech patterns to identify disorders, while other tools help assess children's speech.
Physical-acoustic analysis, which looks at the physical properties of speech itself, includes methods like studying its acoustics, how it's produced, and in-depth examinations of speech affected by neurological disorders. By using this combination of methods, professionals can accurately diagnose and treat speech disorders, ultimately improving communication and quality of life for those affected.

2.1 AI Techniques Used to Detect Impact speech

2.1.1 Dataset used

The dataset used for speech characteristics plays a crucial role in developing AI systems that can accurately recognize and understand human speech.This dataset consists of audio files and transcripts from human conversations, which are used to train machine learning algorithms for speech recognition.

2.1.1.1 Description

The dataset utilized for impact speech detection comprises a collection of audio recordings captured from individuals under various conditions. These recordings encompass instances of speech impacted by factors such as emotion, stress, physical exertion, or environmental noise. The primary goal of the dataset is to facilitate the development and evaluation of machine learning models for identifying and classifying instances of impact speech.

2.1.1.2 Characteristics of speech

1. **Dataset Name:** UXTD and UPX
2. **Format:** The dataset is structured in tabular format, with columns representing different attributes or features of the data. Each row represents a unique instance or sample within the dataset[31]. The columns include:

Tableau 2.1: A dataset of audio recordings from typically developing children

columns	attributes or features of the data
UXTD	**Ultrax Typically Developing Children**
Speaker_id	Unique identifier for each speaker.
Sex	Gender of the speaker (Male/Female).
Age_y	Age of the speaker in years.
Age_m	Age of the speaker in months.
Age	Combined age in years and months (e.g., 11.83).
Group	Group designation (e.g., 1 for primary school children).
Test_date	Date when the speech test was conducted.
Subset	Subset designation (e.g., train, test, dev).

Tableau 2.2: A dataset of audio recordings from children with speech sound disorders.

columns	attributes or features of the data
UPX	**UltraPhonix**
Dataset	The dataset identifier.
Download Subset	The subset of the dataset from which the data was downloaded.
Speaker_id	Unique identifier for each speaker.
Sex	Gender of the speaker (Male/Female).
Age_y	Age of the speaker in years.
Age_m	Age of the speaker in months.
Age	Combined age in years and months (e.g., 11.83).
SSD Subtype	Type of speech sound disorder the speaker has.

2.1.2 Physical-acoustic speech

Physical-acoustic features refer to measurable attributes derived from the acoustic properties of speech signals (Fig 2.1), including pitch, intensity, spectral characteristics, and temporal dynamics. These features play a crucial role in impact speech detection by capturing changes in vocal production associated with emotional arousal, stress, or physiological exertion. Methods for extracting and processing physical-acoustic features involve preprocessing raw speech signals, extracting relevant features using signal processing techniques, and integrating them into machine learning models for impact speech detection tasks [44].

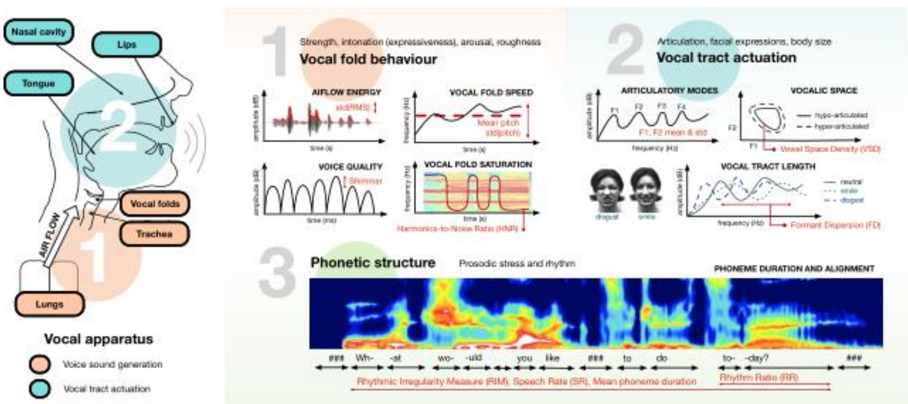

Figure 2.1: Anatomical voice production mechanisms and corresponding acoustic features.

2.2 AI Techniques (algorithms)

2.2.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) have emerged as a powerful class of deep learning models, particularly effective in handling tasks related to image and audio and signal processing. In recent years, CNNs have gained significant attention due to their remarkable performance in various applications, including computer vision, speech recognition, and natural language processing.

2.2.1.1 Definition (CNNs)

CNNs are commonly used in image recognition tasks, but it have been widely used in speech processing tasks [50], including speaker recognition, emotion recognition, and speech recognition [36]. They are particularly effective in capturing temporal variations in speech signals, which are critical for impact speech detection .

2.2.1.2 Architecture (CNNs)

The CNN has several hierarchy levels composed of routing layers and grouping layers, which are defined by a broad variety of charts. In general, CNN begins with a convolutionary layer that accepts input level data. For convolutionary operations with few filter maps of the same dimension, the convolution layer is liable. In addition, the output from this layer is transferred to the sample layer that decreases the scale of the next layers. CNN is locally related to a vast variety of deep learning techniques. These networks are then implemented on the basis of GPU architecture on a number of hundred cores. The role maps will be allocated on the basis of the previous layer knowledge blocks [42].

It depends on the dimensions of the maps. However, each thread is bound to a single neuron by means of a single block of many threads. Similarly, neuron convolution, induction, and summation are carried out over the remainder of the method. Finally, a global memory stores the performance of the above processes. A backward and propagation model is adopted for the efficient processing of results. However, a single spread would not yield positive outcomes, so pulling or moving operations contribute to parallel spread. In addition, the neurons of a single layer interact with a separate number of neurons, influencing the boundary effects[53].

In Figure 1, the general architecture of CNN is explaining the work of this deep learning neural network. A deep learning algorithm includes input preprocessing, deep learning model training, storage of the learned model, and the last phase of the model implementation. In these phases, the most computational (or data intensive activity is to train the deep learning algorithms (defining and running). The model is provided some input through a neural network that produces some output at the specified step (also called forward transmission).

The weights are changed if the performance is inappropriate or inaccurate (backward pass). This could be like a basic matrix multiplication, where input (first matrix row) for such unique output objects is multiplied by weight (second matrix column). Serial systems (CPU-based) are typically not feasible for higher order matrices (large inputs

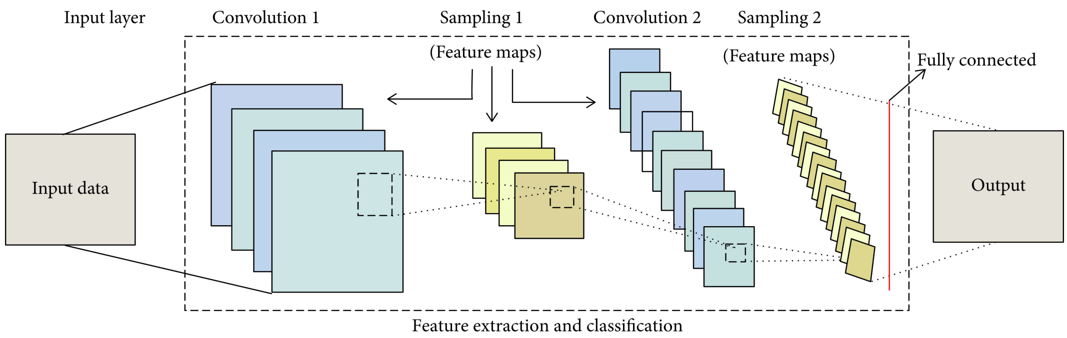

Figure 2.2: Architecture of CNN [42]

and weighs). Fortunately, GPU delivers much superior options than conventional single or cluster CPU systems [43] of graphic processing units for general purposes.

2.2.2 Feature Extraction Methods like MFCC

Mel Frequency Cepstral Coefficients (MFCC) utilize a filter bank for signal analysis, transforming the signal's short-term energy into a series of coefficients through a discrete cosine transform (DCT). This process is based on the logarithmic scale of Mel frequencies. Widely applied in various speech and voice recognition systems [55] , MFCCs are instrumental in tasks such as confirming airline reservations and authenticating spoken digits over the phone for security verification. Enhancements to the foundational MFCC technique have been suggested to increase its resilience, including elevating the logarithmic Mel amplitudes to a power near 2 or 3 prior to the DCT application. This adjustment aims to diminish the influence of segments with lower energy levels[49]

2.2.3 Transfer Learning

Transfer learning involves using knowledge gained from solving one problem and applying it to a different, but related, problem [58]. In the context of speech detection, transfer learning can be particularly useful for leveraging pre-trained models on large datasets and adapting them to smaller datasets specific to impact speech detection tasks. By fine-tuning pre-trained models on smaller datasets, transfer learning can improve the performance of impact speech detection systems, especially when labeled data is limited.

2.2.3.1 Pre-trained Models

Pre-trained models, such as those trained on large-scale speech recognition datasets like LibriSpeech or CommonVoice [5], provide a starting point for transfer learning in impact speech detection tasks. These models have learned general features from vast amounts of data, which can be adapted to specific tasks with smaller datasets.

2.2.3.2 Fine-tuning Techniques

Fine-tuning involves updating the parameters of pre-trained models using a smaller dataset specific to the target task [41]. Techniques such as freezing certain layers of the pre-trained model and training only the remaining layers on the target dataset can speed up convergence and prevent overfitting.

2.3 Speech disorders Detect Technique:

Detecting speech disorders involves a range of methods often using technology to analyze how people speak. One effective way is, by employing learning models to identify speech signals that indicate a disorder. These models can assess aspects of speech, such as vowels, consonants and sentences to distinguish between disordered speech[57]. Another technique involves tools for evaluating speech issues in children in a manner. These tools can offer assessments compared to evaluations done by humans and are beneficial for professionals like speech language pathologists, pediatricians and teachers who work with young children[3]. For a examination features like mel frequency cepstral coefficients (MFCC) and i vectors are utilized to pinpoint specific language impairments (SLI) and differentiate them from mispronunciations due to using a second language. Statistical tests such as ANOVA are used to identify the features, which are then fed into pretrained models like VGG 16, MobileNet v2, ResNet 50 and ResNet 101 for diagnosis[3].

2.4 Others methods for physical-acoustic speech:

Physical-acoustic speech analysis is a critical aspect of speech disorder assessment and treatment. Here are some methods used in this field:

- **Acoustic Phonetics:** This method focuses on the acoustic properties of human speech, such as frequency, duration, pitch, and intensity. Tools like PRAAT software are used to visualize, play, annotate, and analyze sound objects based on these properties. For example, in assessing speech impairments caused by dysarthria, acoustic phonetics can analyze voice onset time (VOT) and fundamental frequency (F0) to complement perceptual-based assessments[38].
- **Experimental Phonetics:** This area investigates the physical instantiation of speech by studying how speech is produced, perceived, and the nature of the acoustic signal transmitted from speaker to listener. It addresses questions about phonetic structures of specific languages and the cognitive-physical relation of speech instantiation [33].
- **Voice Assessment:** Updates in voice assessment include a variety of methods such as perceptual, acoustic, aerodynamic, and endoscopic imaging. These

methods provide a comprehensive approach to understanding and diagnosing speech disorders[4].

- **Acoustic Studies of Dysarthric Speech:** These studies can be deterministic or stochastic and focus on phonetic aspects, physiological components, or both. They involve measurements for various aspects of phonation, articulation, and resonance, which are crucial in the assessment of neurologically disordered speechs[48].

Conclusion

there are various methods and techniques used for physical-acoustic speech analysis. These methods include spectrographic analysis, formant analysis, voice range profile, voice onset time analysis, and fundamental frequency analysis. These techniques allow researchers and professionals to analyze the acoustic properties of speech sounds, measure vocal capabilities, assess vocal health, and study aspects of speech production such as vowel quality, pitch variation, and articulatory characteristics. By utilizing these methods, researchers and professionals can gain valuable insights into speech production and perception, which can contribute to the understanding and treatment of speech disorders, language development, and other related areas of study.

CHAPTER 3:

Proposed Approach And Experimental Results

Chapitre 3

Proposed Approach And Experimental Results

Introduction

This chapter presents the experimental results of applying machine learning to audio data analysis, focusing on speech sound disorders (SSDs) in children. It details the use of essential libraries (NumPy, Librosa, TensorFlow) and Google Colab's computational resources. The implementation covers environment setup, CNN model architecture, and training visualization tools, analyzing metrics like model accuracy, loss, and category performance, supported by visualizations and a confusion matrix. The chapter concludes with an F1 score analysis and highlights advances in SSD detection and intervention. Additionally, it discusses the subjective nature of current SSD assessments and the potential of machine learning to automate these processes. The chapter introduces UltraSuite, a curated repository of synchronized ultrasound and acoustic data from child speech therapy sessions, aiming to facilitate further research in SSD detection and intervention.

3.1 Proposed Dataset:

Based on the search results, we found a dataset relevant to the detection and intervention of speech disorders:

3.1.1 Using Data:

The data was extracted from an information repository known as "Ultrasuite," which is a fundamental and significant source containing a diverse range of audio data from speech therapy sessions for children, including ultrasound and acoustic data. The current release comprises three datasets, one tailored for typically developing children and two for children with speech disorders. Additionally, "Ultrasuite" includes a set of comments, some manually produced and some generated automatically, and provides tools for processing, transforming, and visualizing the data.
Two datasets were selected:

- **- Ultrax Typically Developing - UXTD:** This dataset includes 58 typically developing children.
- **- UltraPhonics - UPX:** This is a secondary dataset comprising children with speech disorders, with data collected from 20 children.

3.1.2 Proposed technique:

One proposed technique involves the application of a Convolutional Neural Network (CNN) to classify patients and healthy individuals, as well as identify the type of disease from audio and ultrasound data. The dataset consists of recordings capturing speech sound disorders (SSDs) and the development of speech skills in children. The data is categorized into two main classes:

- UXTD: Represents typically developing children.
- UPX: Represents additional children suffering from speech sound disorders.

To implement this technique:

3.1.2.1 Data Preprocessing:

Begin by preprocessing the data, including extracting relevant features such as spectrograms from audio recordings. Normalize the data to ensure consistency across samples and split it into training, validation, and test sets to facilitate model training and evaluation.

3.1.2.2 CNN Model Architecture:

Design a CNN architecture suitable for processing spectrogram features. This architecture typically includes convolutional layers to capture spatial patterns, activation functions like ReLU for introducing non-linearity, pooling layers for downsampling, and fully connected layers for classification using softmax activation.

3.1.2.3 Mel-Frequency Cepstral Coefficients (MFCC):

MFCC, or Mel-Frequency Cepstral Coefficients, is a common technique for extracting features from audio signals and is widely used in speech signal processing [24]. MFCC is important because it mimics the way the human ear perceives sound, making it effective in analyzing and classifying sound characteristics.
MFCC is used to convert the audio signal into a set of features that express the spectral characteristics of the sound. This is done by analyzing the audio signal into short frames and computing the cepstral spectrum for each frame. Then, the Mel frequency scale is applied to convert the spectrum into frequency scales that correspond to human ear perception. Subsequently, the logarithms of the energies in these scales are computed and transformed back into the time domain using Inverse Discrete Cosine Transform (IDCT).

3.1.2.4 Model Training:

Initialize and train the CNN model on the training dataset using optimization techniques like Adam or SGD. Validate the model's performance on the validation set to prevent overfitting, utilizing techniques such as dropout regularization if necessary.

3.1.2.5 Evaluation and Fine-tuning:

Evaluate the trained model's performance on the test set, adjusting hyperparameters and model architecture based on evaluation results to enhance classification performance.

3.1.2.6 Inference:

Deploy the trained CNN model to classify new audio samples into the appropriate categories, i.e., patients or healthy individuals, and identify the type of disease. Monitor the model's performance in real-world scenarios, refining it as needed to ensure accurate classifications.

3.1.3 Metrics:

When evaluating speech disorder detection and intervention, there are several metrics that can be used to assess the effectiveness of these processes. Here are some key metrics to consider:

3.1.3.1 Accuracy:

Accuracy is a measure of classification model performance used to assess the extent of correctness in the model's classifications. Accuracy is calculated by dividing the number of correctly classified items by the total number of items in the dataset.
The mathematical formula for calculating accuracy is:

$$\text{Accuracy} = \frac{\text{Number of Correct Classifications}}{\text{Total Number of Classifications}} \times 100$$

3.1.3.2 F1 score:

F1 score is one of the most important machine learning metrics, especially when working on classification models. This indicator is a harmonic mean between precision and recall. Precision is the number of instances actually positive among the positive classifications. It is calculated as the number of true positives divided by the total number of positive predictions.
Mathematically, the F1 score is calculated as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.3}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.4}$$

This metric is particularly useful when dealing with imbalanced datasets, where the accuracy alone might not reflect the true performance of a model. A high F1 score indicates that the model has low false positives and low false negatives, making it a more reliable indicator of the model's predictive power [14] [13] [30].

3.1.3.3 Mean squared error (MSE):

Mean Squared Error (MSE) is a widely used metric in machine learning, especially for regression tasks. It measures the average of the squares of the errors, which are the differences between the predicted values and the actual values. The MSE is calculated using the formula:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

where:
- n is the number of observations,
- Y_i is the actual value, and
- $\hat{Y}_i$ is the predicted value.
A lower MSE indicates a better fit of the model to the data. It's important to note that MSE can be sensitive to outliers, as it squares the errors, thus giving more weight to larger errors. This metric is helpful for comparing different models and for optimizing models during the training process [22] [23].

3.1.4 Confusion matrix:

A confusion matrix is a table used to evaluate the performance of a classification model by comparing the model's predictions with the true values. It is often used in machine learning to assess the quality of a binary classification model (predicting two classes) or multi-class model (predicting more than two classes). The confusion matrix is made up of four elements: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

Tableau 3.1: Confusion Matrix

	Actually Positive (1)	**Actually Negative (0)**
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

The elements of the confusion matrix are used to calculate several performance metrics, such as accuracy, precision, recall, F1 score, and area under the ROC curve.

- **True Positives (TP):** The number of times the model correctly predicted the positive class (true positive).
- **False Positives (FP):** The number of times the model incorrectly predicted the positive class (false positive).
- **True Negatives (TN):** The number of times the model correctly predicted the negative class (true negative).
- **False Negatives (FN):** The number of times the model incorrectly predicted the negative class (false negative).

3.2 Environment and Development Tools

In the realm of data-driven applications and scientific computing, Python stands out as a versatile and powerful programming language. Its popularity stems from its simplicity, readability, and extensive ecosystem of libraries tailored for various domains.

3.2.1 Python language

Python is a high-level, interpreted programming language known for its simplicity, readability, and versatility. Created by Guido van Rossum and first released in 1991[35], Python has since gained widespread adoption across various domains, including software development, web development, data analysis, machine learning, scientific computing, and more.

It is characterized by its clean and concise syntax, which emphasizes readability and reduces the cost of program maintenance.Among the most prominent libraries in the Python ecosystem are:

3.2.1.1 Scikit-Learn

Scikit-Learn, often abbreviated as sklearn, is a popular machine learning library in Python that offers an extensive collection of simple and efficient tools for data mining and analysis [54] . It is built on top of NumPy, SciPy, and Matplotlib, making it easy to integrate with other scientific computing libraries in the Python ecosystem. Scikit-Learn provides implementations of various machine learning algorithms, including classification, regression, clustering, dimensionality reduction, and model selection. Additionally, it offers utilities for preprocessing data, such as feature scaling, feature selection, and data transformation. One of the key strengths of Scikit-Learn is its user-friendly API, which follows a consistent interface across different algorithms, making it easy for users to experiment with different models and techniques. Moreover, Scikit-Learn emphasizes code simplicity and readability, making it an excellent choice for both beginners and experienced practitioners in the field of machine learning.

3.2.1.2 Pandas

Pandas is a powerful Python library that has become the go-to tool for data analysts and scientists alike. It is widely adopted in various industries and academic fields due to its versatility and ease of use [51]. At its core, Pandas provides data structures called DataFrames, which are essentially two-dimensional labeled arrays. These DataFrames allow users to organize, manipulate, and analyze data in a tabular format, similar to working with a spreadsheet or database table.

3.2.1.3 NumPy

NumPy, a foundational package in Python for scientific computing, enables operations on multi-dimensional arrays and matrices [6]. Serving as the basis for numerous scientific computing libraries, NumPy facilitates efficient storage and manipulation of large datasets, with its central data structure, the ndarray, enabling swift and efficient mathematical computations on arrays. Additionally, NumPy offers an extensive library of mathematical functions that support element-wise operations on arrays, empowering intricate numerical computations and array manipulations, including basic arithmetic operations, trigonometric functions, exponential and logarithmic functions, statistical operations like mean and median, and robust linear algebra capabilities such as matrix multiplication, eigenvalue decomposition, and singular value decomposition.

3.2.1.4 Librosa

The Librosa library in Python is a crucial tool for analyzing audio and music data. It is especially valuable for identifying patterns and similarities within extensive audio datasets by extracting features like chroma pitch, RMS, tempo, beat onset detection, and separating percussive and harmonic elements. Librosa also offers functionalities for modifying the pitch and speed of audio files, enabling comparisons between files with varying keys or tempos[32].

3.2.1.5 TensorFlow

TensorFlow is an open-source, end-to-end platform for machine learning that provides a comprehensive and flexible ecosystem of tools, libraries, and community resources. This ecosystem empowers researchers to advance the state-of-the-art in machine learning and enables developers to easily build and deploy applications powered by machine learning technology[2].

3.3 Computing Resources

Colab is a free Jupyter Notebook environment that runs entirely in the cloud [16] [15] [34]. It is especially well suited to machine learning, data science, and education1. Colab notebooks execute code on Google's cloud servers, meaning you can leverage the power of Google hardware, including GPUs and TPUs, regardless of the power of your machine. All you need is a browser.

3.3.1 Implementation of the proposed method

3.3.2 Working Environment

This code implements a machine learning project using audio data, covering importing, analysis, training, and evaluation.

```python
import glob
import os
import librosa as lib
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import librosa.display
import numpy as np
import IPython.display as ipd
import shutil
import soundfile
from tqdm import tqdm
from sklearn.model_selection import StratifiedShuffleSplit
import shutil
from sklearn.preprocessing import LabelEncoder
from keras.utils import to_categorical
from keras.models import Sequential
from keras.layers import Dense, Dropout, Activation, Flatten
from keras.layers import Conv2D, MaxPooling2D
from keras.optimizers import Adam
from keras.callbacks import ModelCheckpoint
from sklearn.metrics import classification_report, accuracy_score, confusion_matrix
from mlxtend.plotting import plot_confusion_matrix
```

Figure 3.1: Import necessary packages

Based on the provided code, it appears that this function creates a CNN model for handling audio data. The function takes parameters such as audio dimensions and the number of classes, and defines the model architecture including convolutional layers. It is then used in later stages of the training and evaluation process.

```python
#Creating a CNN model
def CNN_model(n_width,n_height,n_channels,n_dropout,n_classes):
    cnn_model = Sequential()

    cnn_model.add(Conv2D(filters=24, kernel_size=(5,5), strides=(1, 1),input_shape=(n_width,n_height,n_channels), activation ='relu'))
    cnn_model.add(MaxPooling2D((4, 2), strides=(4, 2)))

    cnn_model.add(Conv2D(filters=48, kernel_size=(5,5), padding = 'valid', activation ='relu'))
    cnn_model.add(MaxPooling2D((4, 2), strides=(4, 2)))

    cnn_model.add(Conv2D(filters=48, kernel_size=(5,5), padding = 'valid', activation ='relu'))

    cnn_model.add(Flatten())
    cnn_model.add(Dropout(rate=n_dropout))

    cnn_model.add(Dense(64, activation ='relu'))
    cnn_model.add(Dropout(rate=n_dropout))

    cnn_model.add(Dense(n_classes, activation ='softmax'))

    return cnn_model
```

```python
cnn_model = CNN_model(x_train.shape[1],x_train.shape[2],x_train.shape[3],0.5,len(encode.classes_))
```

Figure 3.2: CNN Model

This code generates visualizations to analyze the performance of a machine learning model during training. It plots the training and validation accuracy and loss over the training iterations, providing insights into the model's behavior and potential issues like overfitting or underfitting. The information gathered from these visualizations can then be used to improve the model's design and produce better results.

```python
history = cnn_history.history

#Plotting epoch vs Training and Testing accuracy Graph
plt.figure(figsize=(16,6))
plt.plot(history['accuracy'], color = "b")
plt.plot(history['val_accuracy'], color = 'r')
plt.title('Model accuracy')
plt.ylabel('Accuracy')
plt.xlabel('Epoch')
plt.legend(['Training Accuracy','Testing Accuracy'],loc='upper left')


#Plotting epoch vs Training and Testing loss Graph
plt.figure(figsize=(16,8))
plt.plot(history['loss'], color = "b")
plt.plot(history['val_loss'], color="r")
plt.title('Model loss')
plt.ylabel('Loss')
plt.xlabel('Epoch')
plt.legend(['Training Loss','Testing Loss'],loc='upper right')

plt.show()
```

Figure 3.3: Plot Accuracy and Loss Curves

This code analyzes the performance of a trained deep learning model by plotting graphs of accuracy and loss for the training and testing data against the number of training epochs. The code uses large plot dimensions and clear axis and legend labels, making it easy to understand and interpret the model's performance during training and evaluation.

3.3.3 Experiments and results

After training the model for several epochs, we got good results, and these are some pictures showing the model's performance:

3.3.3.1 Model accuracy

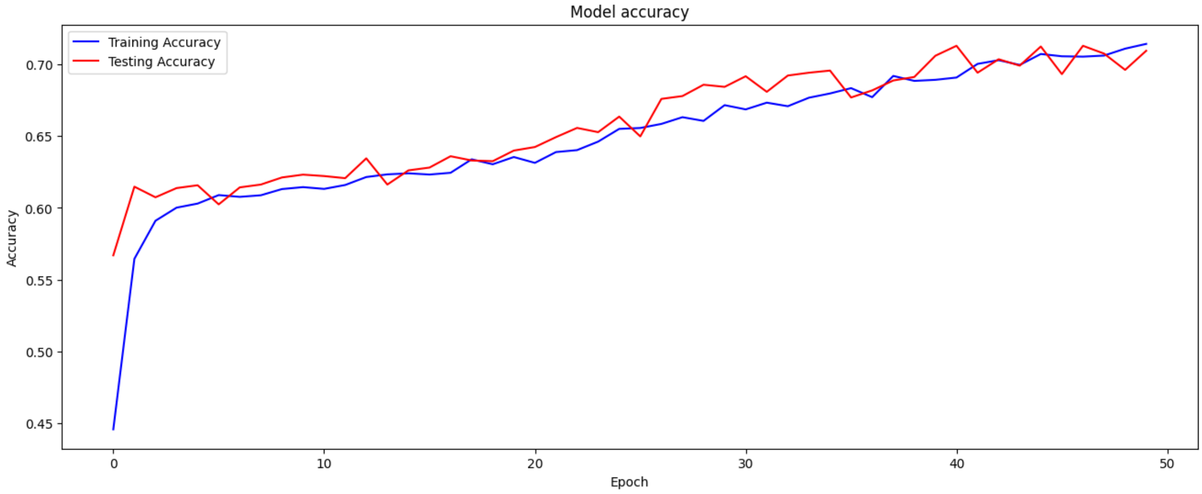

Figure 3.4: Evolution of segmentation accuracy using the CNN model

As you can see in the above figure (3.4), the accuracy curve of both the training and Testing sets are increasing well and at the end they reach excellent accuracy ”Training Accuracy: 0.7823 and Testing Accuracy: 0.7094”

3.3.3.2 Model loss

The figure (3.5) shows the decrease in the cost function over the training epochs of the CNN model:

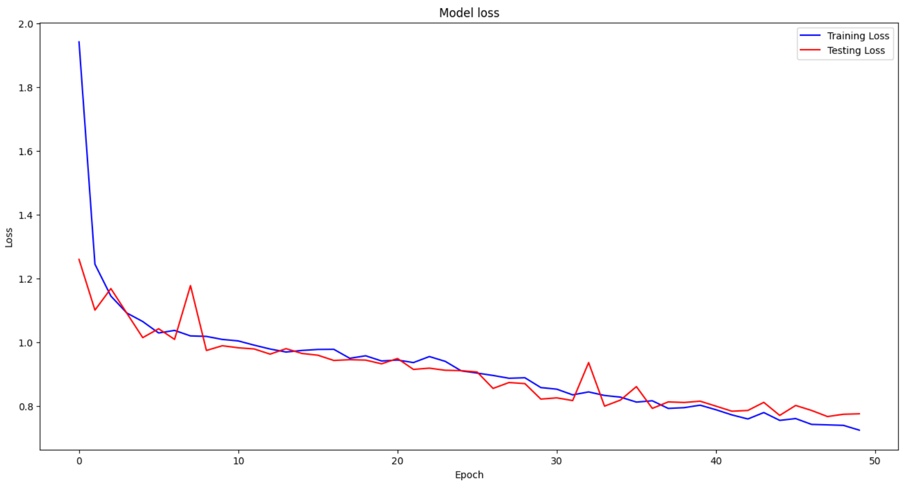

Figure 3.5: Loss function minimization by the CNN model.

Also in the above figure (3.5), the loss function curve of both the training and Texting sets are decreasing well and at the end they reach " Training Loss: 0.6081 and Testing Loss: 0.7756 "

3.3.3.3 Confusion matrix:

Based on the confusion matrix presented in the figure, it can be observed that the model exhibits good performance in classifying some categories while facing difficulties with others.

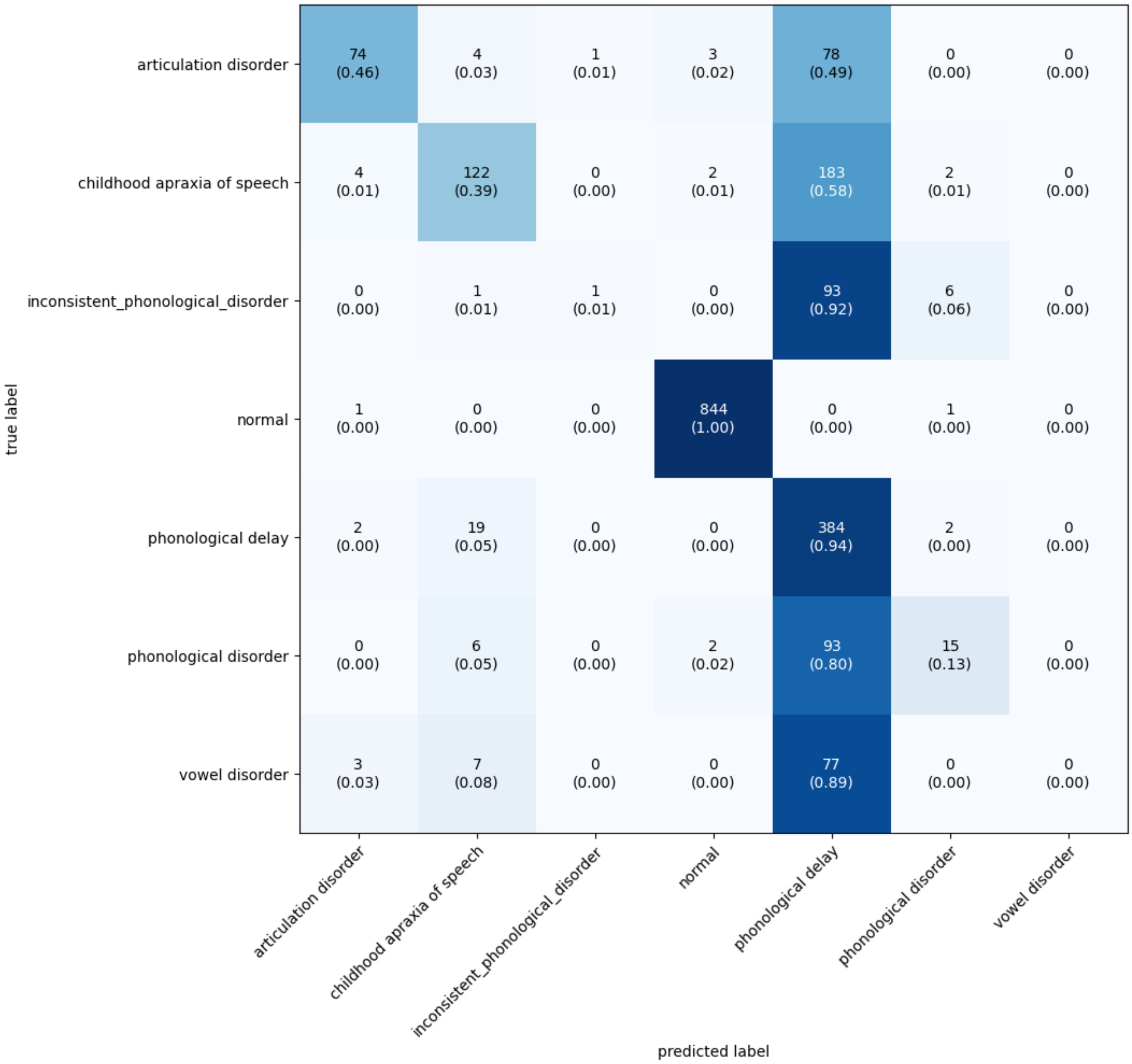

Figure 3.6: Confusion matrix for classifying speech disorders resulting from CNN-based segmentation.

3.3.4 F1 score:

- The np.argmax function is used to get the index of the maximum value, which is helpful for converting the one-hot encoded labels back to their original class labels.
- Class D has no instances, which is why its metrics are zero.
- Class E has very low precision and recall, indicating that the model struggles with this class.
- Overall accuracy is quite high at 0.87, but the macro average F1-score is lower, indicating that the performance varies significantly across different classes.

```
64/64 [==============================] - 11s 164ms/step
                                 precision    recall  f1-score   support

           articulation disorder       0.88      0.46      0.61       160
      childhood apraxia of speech       0.77      0.39      0.52       313
inconsistent_phonological_disorder       0.50      0.01      0.02       101
                          normal       0.99      1.00      0.99       846
              phonological delay       0.42      0.94      0.58       407
           phonological disorder       0.58      0.13      0.21       116
                  vowel disorder       1.00      0.00      0.00        87

                        accuracy                           0.71      2030
                       macro avg       0.73      0.42      0.42      2030
                    weighted avg       0.79      0.71      0.67      2030
```

Figure 3.7: Overall F1 score on the CNN model.

Conclusion

Significant developments in speech disorder detection and intervention have revolutionized the field, particularly for children. Advances such as automated tools, innovative features, mobile applications, specialized software, and standardized assessment methods have enhanced the accuracy, efficiency, and effectiveness of diagnosing and treating speech disorders. Key metrics like accuracy, sensitivity, specificity, precision, recall, F1 score, and Cohen's Kappa are crucial for evaluating these systems. Over time, AI, machine learning, and natural language processing have significantly improved these metrics, resulting in better outcomes and quality of life for individuals with speech impairments.

General Conclusion

Based on the search results for "Recent advancements in speech disorder detection and intervention," it is evident that there have been significant developments in the field of speech therapy. These advancements include the use of automatic speech assessment tools, technological innovations such as teletherapy and virtual reality, increased awareness about the importance of early intervention, and government support for speech therapy services. However, there are also challenges facing the speech therapy market, such as a shortage of qualified speech therapists, a lack of standardized assessments and treatments, and the need to increase awareness about the importance of speech therapy services.

3.4 Advancements in Speech Assessment Tools:

Automatic speech assessment tools have gained popularity among pediatricians and teachers who work with preschoolers. These tools have shown to be more accurate in detecting speech sound disorders than human-based assessment methods [57].

3.5 Technological Innovations:

- Teletherapy, which uses online platforms to provide speech therapy services, has become a significant development in the field. It offers convenience, affordability, and effectiveness, especially for patients who cannot travel to clinics or live in remote areas [18].
- Speech recognition software has become more advanced, making it easier to diagnose and treat speech disorders [18].
- Virtual reality technology has been used to create immersive environments that help patients overcome communication difficulties [18].

3.6 Government Support and Rising Geriatric Population:

There is a growing awareness of the importance of early intervention and diagnosis of communication disorders. Early intervention can significantly improve outcomes for children with speech and language disorders and help prevent more serious difficulties later in life [18].

3.7 Technological Innovations:

- Governments around the world are recognizing the importance of speech therapy services and increasing their support for the industry. This includes mandates for schools to provide speech therapy services to children with communication disorders and the provision of speech therapy services by national healthcare systems [18].
- The rising geriatric population, which is more prone to speech and language disorders, is expected to drive the demand for speech therapy services. Conditions such as dementia and Parkinson's disease are expected to increase, further emphasizing the need for speech therapy [18].

3.8 Technological Innovations:

- There is a shortage of qualified speech therapists in many regions, which can limit access to services [18].
- The lack of standardized assessments and treatments makes it challenging to compare outcomes and determine the most effective treatments for different patients [18].
- Stigma surrounding communication disorders can lead to delays in diagnosis and treatment, negatively impacting outcomes [18].

In conclusion, recent advancements in speech disorder detection and intervention have brought about significant improvements in the field of speech therapy. These advancements include the use of automatic speech assessment tools, technological innovations such as teletherapy and virtual reality, increased awareness about the importance of early intervention, and government support for speech therapy services. However, challenges such as a shortage of qualified speech therapists, a lack of standardized assessments and treatments, and the need to increase awareness about the importance of speech therapy services still exist.

3.9 FUTURE WORK

As research and innovation in artificial intelligence (AI) progress, this approach can be extended to other areas beyond speech disorders. There are promising opportunities to use advanced AI models for early detection of other diseases such as developmental and mental health disorders, using health and behavioral data. Systems can also be developed to monitor environmental exposure to pollutants and their association with health outcomes, and to understand the impacts of pollution and climate change on public health. By integrating data from multiple sources and developing advanced predictive models, these solutions can be expanded to cover underrepresented areas and populations, helping to improve public health in diverse communities.
With further research and development, this approach can be extended to other crops and environmental conditions.

Bibliographie

[1] “4 tips to support your child if they have a speech disorder”.

[2] “About tensorflow”.

[3] “Acoustic speech markers for schizophrenia-spectrum disorders: a diagnostic and symptom-recognition tool”.

[4] “Acoustic studies of dysarthric speech: Methods, progress, and potential”.

[5] “Advancing speech recognition with transfer learning techniques”.

[6] “Array programming with numpy”.

[7] “Chapter iii - assessment methods for young children with communications disorders - continued”.

[8] “Child speech and language”.

[9] “Childhood apraxia of speech”.

[10] “Childhood spoken language disorders”.

[11] “Early identification of speech-language delays and disorders”.

[12] “The effectiveness of a cognitive-behavioral counseling program to reduce some speech disorders in linguistically delayed children( )”.

[13] “F1 score in machine learning”.

[14] “F1 score in machine learning: Intro calculation”.

[15] “Google colab - what is google colab?”.

[16] “Google colaboratory”.

[17] “High-level cnn and machine learning methods for speaker recognition”.

[18] “Insights into the latest trends, technologies, and growth drivers of the speech therapy industry”.

[19] “Language and speech disorders in children”.

[20] “Language and speech disorders in children”.

[21] “Language disorders in children”.

[22] “Machine learning: an introduction to mean squared error and regression lines”.

[23] “Mean squared error (mse)”.

[24] “Mfcc in audio signal processing for voice disorder: a review”.

[25] “Quick statistics about voice, speech, language”.

[26] “Speech acoustics: How much science?”.

[27] “Speech and language disorders in children: Implications for the social security administration’s supplemental security income program.”.

[28] “Speech disorders in primary school students in jeddah”.

[29] “Speech-language pathologists’ practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders”.

[30] “Summary: Understanding the f1 score in machine learning”.

[31] "Ultrax typically developing children".

[32] "Using librosa to change the pitch of an audio file".

[33] "Voice assessment: updates on perceptual, acoustic, aerodynamic,and endoscopic imaging methods".

[34] "What is google colab?".

[35] "What is python? executive summary".

[36] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, L. Deng, G. Penn et D. Yu – "Convolutional neural networks for speech recognition", *IEEE/ACM Transactions on audio, speech, and language processing* **22** (2014), no. 10, p. 1533–1545.

[37] E. A. Beccaluva, F. Catania, F. Arosio et F. Garzotto – "Predicting developmental language disorders using artificial intelligence and a speech data analysis tool", *Human–Computer Interaction* **39** (2024), no. 1-2, p. 8–42.

[38] P. S. Beddor – "503Experimental Phonetics", The Oxford Handbook of Linguistic Analysis, Oxford University Press, 02 2015.

[39] D. V. Bishop – "What causes specific language impairment in children?", *Current directions in psychological science* **15** (2006), no. 5, p. 217–221.

[40] R. M. S. Hala Ibrahim Al-Jarwani – *Stuttering disorders (diagnostic and therapeutic view)*, University Knowledge House, 2013.

[41] J. Howard et S. Ruder – "Universal language model fine-tuning for text classification", *arXiv preprint arXiv:1801.06146* (2018).

[42] B. Jan, H. Farman, M. Khan, M. Imran, I. U. Islam, A. Ahmad, S. Ali et G. Jeon – "Deep learning in big data analytics: a comparative study", *Computers & Electrical Engineering* **75** (2019), p. 275–287.

[43] B. Jan, F. G. Khan, B. Montrucchio, A. T. Chronopoulos, S. Shamshirband et A. N. Khan – "Introducing tope-fft: An opencl-based fft library targeting gpus", *Concurrency and Computation: Practice and Experience* **29** (2017), no. 21, p. e4256.

[44] Q. Jin, C. Li, S. Chen et H. Wu – "Speech emotion recognition with acoustic ahttps://www.activeloop.ai/resources/glossary/mean-squared-error-mse/nd lexical features", *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE, 2015, p. 4749–4753.

[45] C. J. Johnson, J. H. Beitchman et E. Brownlie – "Twenty-year follow-up of children with and without speech-language impairments: Family, educational, occupational, and quality of life outcomes", (2010).

[46] v. d. D. J. V. G. A. J. D. V. M. N. V. H. S. V. M.-A. Kappen, Mitchel – *Acoustic speech features in social comparison: how stress impacts the way you sound*, 2022.

[47] M. K. Karia – "Using acoustic phonetics in the assessment and treatment of speech disorders", p. 441–451, Springer International Publishing, Cham, 2023.

[48] J. P. Kasture, Neha – "Automatic recognition of disordered childrenâ€™s speech signal in dyadic interaction using deep learning models", (2023).

[49] P. K. Kurzekar, R. R. Deshmukh, V. B. Waghmare et P. P. Shrishrimal – "A comparative study of feature extraction techniques for speech recognition system", *International Journal of Innovative Research in Science, Engineering and Technology* **3** (2014), no. 12, p. 18006–18016.

[50] Y. LeCun, L. Bottou, Y. Bengio et P. Haffner – "Gradient-based learning applied to document recognition", *Proceedings of the IEEE* **86** (1998), no. 11, p. 2278–2324.

[51] W. McKinney et al. – "Data structures for statistical computing in python.", *SciPy*, vol. 445, 2010, p. 51–56.

[52] Y. S. A.-N. Moussa Mohammad Amayra – *Introduction to communication disorders*, Dar Al-Fikr, 2014.

[53] K. O'shea et R. Nash – "An introduction to convolutional neural networks", *arXiv preprint arXiv:1511.08458* (2015).

[54] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al. – “Scikit-learn: Machine learning in python”, *the Journal of machine Learning research* **12** (2011), p. 2825–2830.

[55] K. Ravikumar, B. Reddy, R. Rajagopal et H. Nagaraj – “Automatic detection of syllable repetition in read speech for objective assessment of stuttered disfluencies”, *International Journal of Electrical and Computer Engineering* **2** (2008), no. 10, p. 2142–2145.

[56] A. Sansavini, A. Guarini, S. Savini, S. Broccoli, L. Justice, R. Alessandroni et G. Faldella – “Longitudinal trajectories of gestural and linguistic abilities in very preterm infants in the second year of life”, *Neuropsychologia* **49** (2011), no. 13, p. 3677–3688.

[57] G. P. Usha et J. S. R. Alex – “Speech assessment tool methods for speech impaired children: a systematic literature review on the state-of-the-art in speech impairment analysis”, *Multimedia Tools and Applications* **82** (2023), no. 22, p. 35021–35058.

[58] K. Weiss, T. M. Khoshgoftaar et D. Wang – “A survey of transfer learning”, *Journal of Big data* **3** (2016), p. 1–40.

[59] I. A.-A. Zureikat – *Speech and language disorders (diagnosis and treatment)*, Dar Al-Fikr, 2007.


الحمد لله على كل حال
١٢٦٤