ENLIGHTEN DS: ADVANCED TECHNOLOGIES FOR SKILL ENHANCEMENT AND TALENT RECOGNITION IN CHILDREN WITH DOWN SYNDROME

Jayasuriya S.H

IT21293030

Dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Hons) in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

April 2025

ENLIGHTEN DS: ADVANCED TECHNOLOGIES FOR SKILL ENHANCEMENT AND TALENT RECOGNITION IN CHILDREN WITH DOWN SYNDROME

Jayasuriya S.H.

IT21293030

BSc (Hons) degree in Information Technology Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

April 2025

DECLARATION

I declare that this is my own work and this dissertation1 does not incorporate without acknowledgement any material previously submitted for a degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology, the nonexclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Student Name	Student ID	Signature
Jayasuriya S.H.	IT21293030	

The above candidate has carried out research for the bachelor's degree Dissertation under my supervision.

Signature: Date: 2025-04-11

ACKNOWLEDGEMENT

I would like to extend my heartfelt appreciation to all the individuals who supported me in carrying out the 4th year research project.

First of all, I sincerely thank our Research Project supervisor Prof. Samantha Thelijjagoda, who provided invaluable guidance and encouragement to carry this project forward successfully. My special thanks also go to our co-supervisor Dr. Junius Anjana for his continuous support, timely guidance, technical insights, and valuable advice that enabled us to fulfill our research goals more efficiently.

I would also like to acknowledge Mr. Tharindu Sampath whose prior experience with Down Syndrome related projects helped us gain valuable perspective and direction during our research process.

I'm truly appreciative of the collaboration and teamwork shown by all our Research Project group members whose dedication and effort contributed immensely to the success of this project.

I also express my sincere thanks to the Senehasa Research Center coordinators for their assistance and support during the data collection phase. My gratitude extends to the schools we visited, whose staff welcomed us and facilitated the data collection process.

I'm especially thankful to the parents of the children who gave their consent and trusted us to include their children in this research. Their cooperation was essential for gathering meaningful data and I greatly value their support. My appreciation also goes to the families of the children for their participation and encouragement.

Lastly, I'm forever grateful to our parents and friends for their unwavering support, love and belief in us throughout this journey. Their motivation helped make this project a reality and a success.

ABSTRACT

Down Syndrome is a genetic disorder that originates from an additional chromosome 21, leading to developmental delays and intellectual disability. The most prominent of the delays are in speech and language development and usually restrict cognitive and social development. Because of this, the advantages of excellent speech therapy are very paramount for children with Down syndrome. The primary objective of this project is to develop a much-advanced web-based speech therapy application for children between 5 and 15 years with Down syndrome. It integrates the latest technologies, including NLP for speech recognition, which is very accurate and performs real-time error correction, and motion detection via real-time image processing using OpenCV, while machine learning algorithms adapt exercises based on the progress of a person.

Multimedia components, such as fun quizzes and animations, are integrated to keep users motivated by making the application design very appealing and engaging. Therefore, it should provide visuals that can improve pronunciation accuracy, through facial expressions and lip movements. At the same time, it should also have an easy-to-use interface so that even any learner with limited computer experience can get through it easily. It not only addresses the speech challenges unique to DS children but also promotes a fun and effective therapeutic experience that helps improve communication skills, boost self-confidence, and promote better social interactions and academic performance.

Keywords: Machine learning, NLP, OpenCV, Image processing

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Natural Language Processing	NLP
Convolutional neural networks	CNN
Down syndrome	DS
Artificial Intelligence	AI
Valence Aware Dictionary and sentiment Reasoner	VADER
Virtual Reality	VR
Automatic Speech Recognition	ASR
Picture Exchange Communication System	PECS
Augmentative Alternative Communication	AAC

1 INTRODUCTION

Down syndrome is a genetic disorder where there is an extra chromosome 21 [1]. This chromosomal condition is common, occurring in just about 1 in every 700 live births [2].

Women aged 44 years and above are at a greater risk of having a baby with Down syndrome or any chromosomal abnormality [3]. Down syndrome children will generally show varied developmental delays in physical and cognitive aspects, among them the speech and communication difficulties. These challenges may be different and range from mild to severe, whereas mostly they include problems related to speech clarity, language comprehension, and social communication skills.

Four key factors affect speech development

Facial Profile: The smaller mouth and nose and relatively larger tongue constrict space for speech and result in less intelligibility. Small and high palates provide an additional challenge to small-motor actions for speech.

Hearing: Low positioning of the ears can result in otitis media, causing fluid accumulation, ear infections, and hearing loss that subsequently affects speech clarity and development. **Muscle Tone:** Low muscle tone in people leads to an open mouth, low tongue position, and mouth breathing. The speech might become slow and complicates vocalizations needing specific tongue movements for clear articulation, like /k/ and /g/.

Cognitive Development: Although the range of cognitive skills is wide, speech and language are lower overall than cognitive skills. Some children have comprehension skills quite superior to their ability to express themselves. Repetition plays a key role in learning new skills and vocabulary.

Figure 1- Unique physical features

The importance of communication in expressing needs, making friends, and taking part in social and educational activities is extremely high in children with Down syndrome. Because of the nature of the syndrome, children often have difficulty articulating words clearly, understanding complex sentences, and following conversations. These problems may enhance the potential for frustration, social isolation, and problems in academic and nonacademic settings. Therefore, the development of communication skills is of great importance to enhance the quality of life of those with Down syndrome and to help them achieve their full potential

To enhance the communication skills in children with the chromosomal disorder of Down's syndrome, one innovative module has been designed speech and language processing for the improvement of communication skills. The module addresses pronunciation, comprehension, and other general communication skills in children with Down syndrome from the age brackets of 5 to 15 years. [4]

This speech and language processing component is engaging, interactive, and child-specific for a child with Down syndrome. It is powered by advanced speech recognition and processing technologies that give individual feedback and relevant exercise areas in the development of crucial communication skills. A web application has been designed for children with Down's syndrome aged between 5 and 15 years, which brings to life the component of Speech and Language Processing [4]. Enhancing communication skills in children with Down syndrome is the key to greater independence, social integration, and academic success for this population of children. The Speech and Language Processing module has been encapsulated within a user-friendly web application to minimize the communication gap of the child with Down syndrome by providing an innovative solution to one of the major problems of concern. This application provides personalized, interactive, and accessible learning experiences that help children affected by Down syndrome to improve their ability to communicate and reach their full potential.

Maternal Age	Incidence of Down syndrome
20	1 in 2,000
21	1 in 1,700
22	1 in 1,500
23	1 in 1,400
39	1 in 150
40	1 in 100
41	1 in 80
42	1 in 70

Figure 2- Birth incidence of Down syndrome

Source: https://ndss.org/about

1.1 Background and Literature Survey

Research in speech and language processing for children with Down syndrome has a history, and multiple studies and techniques have been proposed. The following is a summarized analysis related to this work, the state of the art, and how previous work has been based on common problems, with special attention to pronunciation enhancement using natural language processing, motion analysis, and machine learning.

Children with Down syndrome typically present with specific delays in their language development that may affect their communicative skills and quality of life [5], [6]. Some research has argued that children with DS benefit from early intervention and focused counterbalances for language. The profile of children with DS exhibits a complex pattern of development with respect to language, mediated by their cognitive abilities, hearing impairments, and motor skills [7].

In children with DS, language development is characterized by a slower rate of vocabulary acquisition and greater reliance on gestures compared to typically developing peers [8]. Indeed, delays in language milestones are often continued into adulthood and include expressional difficulties. Furthermore, syntax and morphology are compromised in children with DS, which are basic constituents for forming grammatically correct sentences [9]. These difficulties call for multidimensional and comprehensive intervention strategies.

Early research in language development in children with DS noted the benefit of encouraging early exposure to language and communication experiences. Rondal and Buckley [5] described the family as a major factor, stating that parents can influence the development of the early language in children through daily interactions with him by being responsive and engaging in communication activities with him. Their research validated the benefit of intensive, family-centered approaches to prepare the parents of children with DS to reinforce the home environment with adequate support for the child learning to speak.

Speech and Language Interventions: Many speech and language interventions have been developed to improve communication skills in children with DS. Such interventions normally seek to develop articulation, vocabulary, and sentence structure [10]. A systematic review of speech, language, and communication interventions for children with DS 0-6 years old showed that early intervention and parent-child interaction therapy are effective in improving language outcomes [8]. There is another study that reflected ways of learning and teaching communication skills in DS children [11].

The most important strategies concern the use of augmentative and alternative modes of communication for these children, such as the presentation of their needs in ways other than speech. The most common forms of augmentative and alternative communication used with DS kids will include picture boards, speech-generating devices, and sign language. The research conducted by Sigafoos et al. [12] showed that AAC interventions could clearly elevate the level of communication skills in children with DS and could be a tool for lessening frustration, which results in more pleasant social interactions.

Key augmentative and alternative communication tools include:

- **Speech-Generating Devices:** This device of voice output communication aid has been shown to improve speech, cognition, and social interaction. Research proves this through studies carried out from 29 children from ages 3 to 12, the fact that SGDs work for the development of communication skills and minimize frustration. [13]
- **Picture Exchange Communication System:** PECS enables children to communicate through the exchange of image cards, promoting language skills and social interaction. Three studies with 28 children have substantiated its benefits and are therefore effective in enhancing communicative skills. [13]
- **Sign Language System:** MAKATON combines speech, signs, and symbols and has been shown to have a positive effect on language development. It has been pointed out that MAKATON improves language skills, as indicated in two studies with 20 children. [13]
- **PCS (Picture Communication Symbols):** This aids in proper visualization and increases the speed of communication. Its effectiveness was proved in two studies comprising 22 participants, thus validating its utility in improving efficiency in communication. [13]

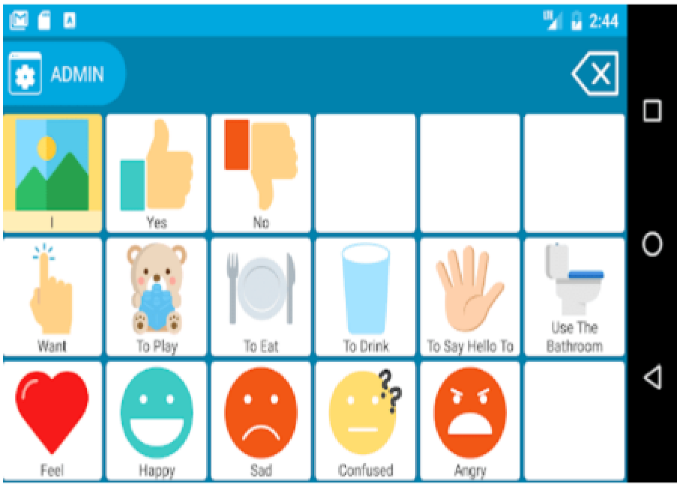

Figure 3-CommBoards AAC Speech Assistance app

Source:https://appsamurai.com/blog/best-apps-for-children-with-down-syndrome/

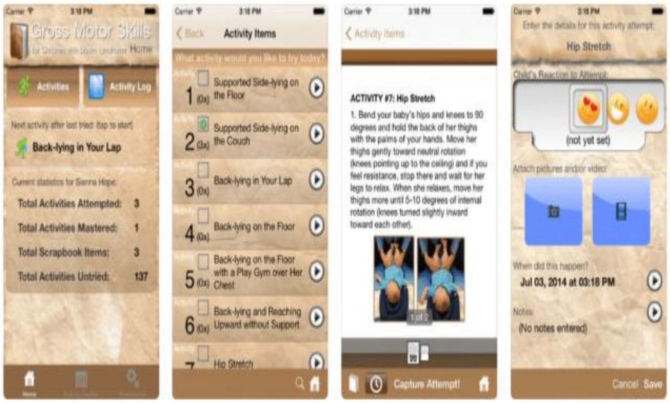

Figure 4-Gross Motor Skills App
Source: https://appsamurai.com/blog/best-apps-for-children-with-down-syndrome/

Intervention programs such as "It Takes to Talk" through the Hannon Center are significant. This program involves training parents how to become their children's best language teachers by using certain strategies in daily activities [14]. Parent-implemented intervention programs have also been effective in significantly improving the language skills of children with DS because they encourage increasingly high-quality interactions.

Recent advances in NLP and ML even open new ways toward the enhancement of the communicative skills of children with DS. Such NLP mechanisms can analyze the patterns in the speech of people and, therefore, give personalized feedback on the improvement of pronunciation [11]. ML algorithms are capable of modeling predictive language development in children with DS for early incidence of language delays and targeted interventions [15].

For instance, Wang and Narayanan [11] tested the feasibility of ASR (Automatic Speech Recognition) systems adapted to children with DS. The authors concluded that when such a system is trained on a set of speech samples collected from children with DS, the system is indeed able to provide speech recognition and intelligent feedback for the speech of a child, which could help correct the articulation right away. These apps can be integrated with mobile infrastructure and hence ease the delivery of services while making them more pleasurable for children with DS.

But in addition to that, ML models will also make use of the examination of large speech recordings dataset to find the common pattern and potential predictors of language development. Hailpern et al. [15] have designed a prediction model having ML algorithms tailored for the prediction of language outcomes stemming early speech and language assessment. These might be the models to be used in assisting clinicians and educators in designing and implementing tailor-fit interventional strategies that will be directed to the exact need of a child.

This can be illustrated by applications like "LENA," which allows the potential to record and analyze speech environmental data from children using wearable technology. These data could be leveraged to provide information and guidance to boost language-rich interactions of caregivers [16]. All this innovative technology enables more data-driven intervention strategies that can be applied to greater effect when intervention becomes increasingly individualized.

Motion analysis, using sensors and cameras to track articulatory gestures, has been harnessed to further refine pronunciation in children with DS [17]. This approach can yield information regarding articulation, which would then lead intervention efforts toward the goal of emission clarity. An intervention study using both motion analysis and NLP-based feedback led to substantial improvement in terms of pronunciation among children with DS [18].

Motion analysis in speech therapy with children with DS comes when they produce words through the movement of their tongues, lips, and jaws. After the movements, the recorded data helps to determine the performance of certain articulations in their normal speech production and deviations from the norm [19]. By offering feedback on these results, visual feedback provides a way in which children can learn to change their movements articulatory patterns to more intelligible speech production.

In a study by Kim and Lee [18], they used motion capture technology in designing a virtual environment for children with DS. The virtual environment was aimed to give the children a practice of how to make the different sounds of speech, with real-time visual feedback of their articulatory movements so that the children can make appropriate adjustments for improved pronunciation. The results indicated that the children improved significantly with the use of the motion analysis system regarding the clarity of their speech in contrast to having received the traditional speech intervention.

Another very promising line of approaches is the application of ultrasound imaging in visualizing one's tongue movements while he or she speaks. The use of this technology is unobtrusive, so real tongue movements can be visualized through it to analyze the necessary coordination that underlies the correct production of speech. Research has shown that ultrasound biofeedback is more effective in children with DS, so all these articulatory adjustments can be shown to children in real time

Speech and language processing in augmenting the communications skills of children with DS are at a state-of-art level through vast research conducted on NLP, ML, and motion analysis approaches. Finally, with respect to future directions, the operation of ML algorithms on a large scale to develop personalized language interventions and integration of motion analysis in the provision of NLP-based feedback deserve further exploration to enhance usage of communication skills among children with DS through VR-based interventions.

Other emerging practices among speech and language therapy involve using virtual reality and augmented reality environments. Such immersive technologies make it possible to create experiences sufficient to engage the child in practicing communication skills. The use of a novel immersive technology will serve to abstract the natural environment but provide support for children with DS to practice conversational skills in a safe and controlled way. It has been shown that early intervention of VR/AR therapy has positive language and social benefits for the children with DS [20]

Another promising line of development is the creation of multi-modal interventions that incorporate a variety of technologies to most effectively address the divergent needs of children with DS. For example, multi-modal systems could integrate NLP, motion analysis, and VR for a complete, personalized language intervention [21]. In such a manner, the systems could attend holistically to speech and language therapy, which deals with speech production, social communication, and cognitive skills.

Furthermore, advances in wearable technology are likely to have a major impact in the years ahead on effective speech and language interventions for children with DS. Wearable technology with the integration of appropriate sensors and function for speech recognition can provide continuous monitoring of a child's spoken language and give feedback instantaneously across typical daily routines [22]. This continuous, in-context feedback could serve to continue to increase the rewarding value of language learning and serve as a reminder that strategies are being applied in a consistent manner.

1.2 Research Gap

This research A [23] investigates the change in language during teenage years in people with Down syndrome. It is based on previous research work conducted by Buckley and Sacks, 1987. In this study, the details of speech and language skills in 90 teenagers having Down syndrome have been studied by the authors. Many adolescents had weak grammatical and syntactic skills; this is often reflected in their "telegraphic" speech. It was very hard to understand them when they spoke, especially during communication with strangers.

Teaching Approaches

- Speech-Only Method (S): Verbal repetition as a means of learning sentence structures with no visual aids.
- Speech-Plus-Reading Method (SR): Students learned through a combination of verbal repetition with reading in which they were shown pictures with sentences written under them.

The SR intervention was most effective and convenient for all children with poor reading skills and very limited memory capacity. Interactivity in book design and use has proven to be enjoyable. Other practical and content-based approaches using visual and physical materials such as Polaroid photographs have shown that application and reinforcement generalize to the everyday spontaneous speech of adolescents.

The focus of the research B [24] is to develop a mobile application that makes use of interactive learning tools and aids in enhancing the cognitive and motor skills of individuals with Down syndrome. The application made use of the Troncoso Method for literacy development, featuring a reading interface where users read small stories; the correct words are shaded to mark progress.

Scrum methodology will be applied along the project, with sprints—review and evaluation of the tasks completed by the team—and retrospectives, where improvements are discussed. Some technologies that will be used are: Kotlin for programming, SQLite for data management, and Balsamiq for UI design.

It relates its findings to other research, like an application called Hatle, developed for children who have reading and writing difficulties. It notes that, while similar technologies have improved cognitive and motor abilities, they have not shown the same effectiveness in selective attention. It also points out difficulties in the education of people with Down syndrome.

This research C [25]describes the design, development, and evaluation of a speech therapy mobile application in children with Down's syndrome. Interactive game for Android using Eclipse with the And Engine library for better handling of the graphics was created. Usability testing of this application has been conducted with the participation of three children between 7 and 10 years who have weaknesses in vocabulary and pronunciation. It applies a systematic speech therapy technique. This technique involves teaching Persian letters, word repetition, and the use of visual and audio aids for pronunciation and practice in speech. The app runs on Android devices with recording capability, extracting features using Mel-Frequency Cepstral Coefficients and using Dynamic Time Warping for speech recognition.

Three main components

- teaching Persian letters.
- practicing pronunciation with feedback.
- game that reinforces learning by having children identify words that contain specific letters.

It improves learning by spotting words containing targeted letters, then allowing children to practice reading. The children's focus was more toward the features of an engaging graphical interface with interactive features of the app, and the children with mild and moderate DS responded better as compared to those who had severe DS and required additional support.

For speech therapy to be effective, it was very effectively integrated with visual and audio elements.

The research D [26] presents the design, development, and preliminary evaluation of a speech therapy software application for treating speech sound difficulties among non-native Arabic-speaking children with Down syndrome at the Rashed Paediatric Therapy Centre in Dubai.

There are two speech recognition algorithms created

- one word-based, where the system considers whole words for grading,
- phone-based, focusing on single phonemes.

Both were implemented using the Nuance Speech Recognition System and Java. The word-based required determination of multiple variants of the word for feedback and the phone-based focused on very specific phonological issues that could give very precise feedback.

The app's interface, based on the educational tool described by "Speaking for Myself," includes printed words, pictures, and audio prompts. While testing the application with the help of the pre-recorded audio, the word-based algorithm showed a 41.1% success rate because of technical and human factors; the phone-based algorithm was more accurate, reaching accuracy of about 73.8%, and rose to 84.2% after fine-tuning of the recognition process.

The outcome of this study underpins the role of highly specialized speech therapy applications that have a great potential to improve the communications skills of the child with Down's syndrome.

EnlightenDS system will be a user-friendly web application, particularly targeted at children aged 5-15, including gamified interactive quizzes for assessment and knowledge enrichment. It will be an interesting and efficient method of learning. The system will also improve pronunciation by practicing exercises that can help children improve their speech.

A significant novelty of the proposed system is that it covers many gaps found in existing research.

- Monitor lip movements during pronunciation exercises to help bring improvement in the accuracy of voice detection in a child.
- The system will get into facial expressions of the child and provide encouraging feedback that shall be personalized. The system shall establish a more supportive and motivating environment to learn, hence more effective speech therapy, by being attuned to the child's emotional state.
- Use new technologies for the implementation.

Features	**Research A**	**Research B**	**Research C**	**Research D**	**EnlightenDS**
Web Application	✗	✗	✗	✗	✓
Monitor children's emotional expressions in real time.	✗	✗	✗	✗	✓
Assess children's knowledge levels using gamified quizzes.	✓	✓	✓	✓	✓
Emotional based pronunciation practice game.	✗	✗	✗	✗	✓
Child voice detection	✗	✓	✓	✓	✓
Use new technologies. (CNN, NLP, OpenCV)	✗	✗	✗	✗	✓

Table 1-Overall Research Gap

1.3 Research Problem

The main research problem for this study is that there is significant delay and challenges in the speech and language development of children with Down syndrome, which is significantly affecting communication skills and reducing the quality of life. Indeed, although receiving many interventions, most children with DS still face articulation and vocabulary acquisition issues, sentence structure problems, all affecting their social and academic interaction.

Voice Detection Accuracy: Technical challenges encountered in the study included Voice Detection Accuracy, as speech recognition algorithms found it difficult to recognize children's voices. Variations of pitch, volume, and clarity, especially in those with Down syndrome, who generally have typical speech patterns and articulation difficulties, created a barrier to correct voice recognition. This is why there is a need to fine-tune the system so that these variations might be accommodated for the effective functioning of the system in speech therapy.

Emotional Expression Challenges: This may have occurred because the application did not fully account for or integrate the emotional states of the children, for example, frustration or motivation. Emotion is a major determinant for the degree of engagement and effectiveness in therapy. If no measures are put in place to recognize and respond to these feelings, the app may miss providing necessary support and feedback that would impact the success of the therapy.

Low Range Auditory Memory: Is when children with Down's syndrome, under the speech-only method of intervention, cannot remember what has been said and thus fail to repeat spoken sentences. Since this technique is based on auditory processing, children who have a little auditory memory fail to reproduce sentences the way they were said, which was inaccurate, hence slow in developing speech.

Difficulty in Sentence Structures: Indicates that the problem is that some children were not able to master complex sentence structures using speech only. This perhaps does not give sufficient support in developing intricate skills of sentence construction, very important for a speaker to convey their ideas across. Therefore, the difficulty about the mastery of sentence structures would immediately and additionally result in less than optimum outcomes in the development of speech for such children.

Selection of the teaching methods adopted: that the study may not have used optimum teaching approaches with those with Down syndrome. If the selected methods were very ill-aligned with the learning peculiarities of the studied population, then the speech characteristics should have minimized its effectiveness. Such a mismatch between the teaching methods and children's needs could have influenced the general results.

Limited Personalized Feedback: Traditional speech therapy methods often do not offer immediate, personalized feedback based on a child's specific speech patterns. More therapeutically adaptive and responsive tools are desperately needed for children with DS. Research shows that individualized feedback further personalizes the impact of speech interventions.

Technology Integration in Interventions: While developments in NLP, ML, and motion analysis look promising, rigorous integration within practical speech therapy tools remains user-friendly. Current technologies are too complex or little accessible for everyday use by children with DS and their careers. Proper integration of these technologies has the potential to revolutionize speech therapy practices.

Engagement and Motivation: Engaging and maintaining the interest of children with DS in speech therapy is a continuous challenge. Traditional interventions can be too simplistic and do not provide the interactivity and fun needed to keep them participating actively. Interactive and engaging tools have been shown to improve adherence to therapy and lead to better outcomes.

1.4 Objectives

1.4.1 Main objective

This project aims to develop an advanced and user-friendly web application to improve pronunciation and general speech development in children with Down syndrome aged 5 to 15 years. The resulting personalized therapeutic interventions for improving communication skills, increasing self-confidence through better socialization, and enhancing academic performance will warrant the use of technologies such as speech recognition, sentiment analysis, motivational messaging, and deep learning.

1.4.2 Sub objectives

Enhance Pronunciation

The application will incorporate advanced speech recognition and deep learning algorithms to improve pronunciation accuracy. Speech will be analyzed at the phoneme and word levels by speech recognition APIs, including Google Cloud Speech-to-Text and Microsoft Azure Speech Service. Using NLP techniques, detailed feedback on articulation and phonological patterns will be provided. Deep learning frameworks like TensorFlow or PyTorch will be used to develop models for fine-tuning toward better pronunciation assessments and corrections.

Sentiment Analysis and Motivational Messaging

Sentiment analysis and motivational messaging will keep the children engaged and motivated. Tools for sentiment analysis, such as VADER or TextBlob in this case, will be utilized to track and make sense of all emotional cues from a user's interactive activities. Convolutional neural networks (CNN) for facial expression and other visual data will be assessed with respect to emotional states. An application, drawing from these analyses, will further proceed to provide personalized motivational messages of encouragement tailored to suit every emotional need of a child in order to promulgate a positive and supportive learning environment.

Motor Skills Tracking and Analysis

In particular, the application shall monitor and analyze motor skills of lip movements. Technologies like OpenCV will be used for the real-time tracking and analysis of lip movements in talking activities. Convolutional Neural Networks (CNN) will be used to process videos in order to increase the accuracy in speech detection. This is a way to provide very fine articulatory skill feedback to children so that they can work on their pronunciation by detailed visual and motion analysis.

Interactive Learning Tools in Practice of Pronunciation

For better pronunciation, the project will incorporate some interactive learning tools that will aid in engaging children. The visual cues will consist of animation or diagrammatic illustrations that clearly show correct pronunciation, including lip and mouth movements. These visual aids will help the children to learn how the sounds are produced and see them more clearly to replicate them. There will also be developed some practices that can be obtained by children while attempting to state the names of things in appropriate contexts, like in sentences and stories. This would serve the purpose of providing practice in context so that there can be reinforcement of the skills learned better for better application of pronunciation skills in everyday communication.

2 METHODOLOGY

2.1 Methodology

2.1.1 Component Specific System Architecture Diagram

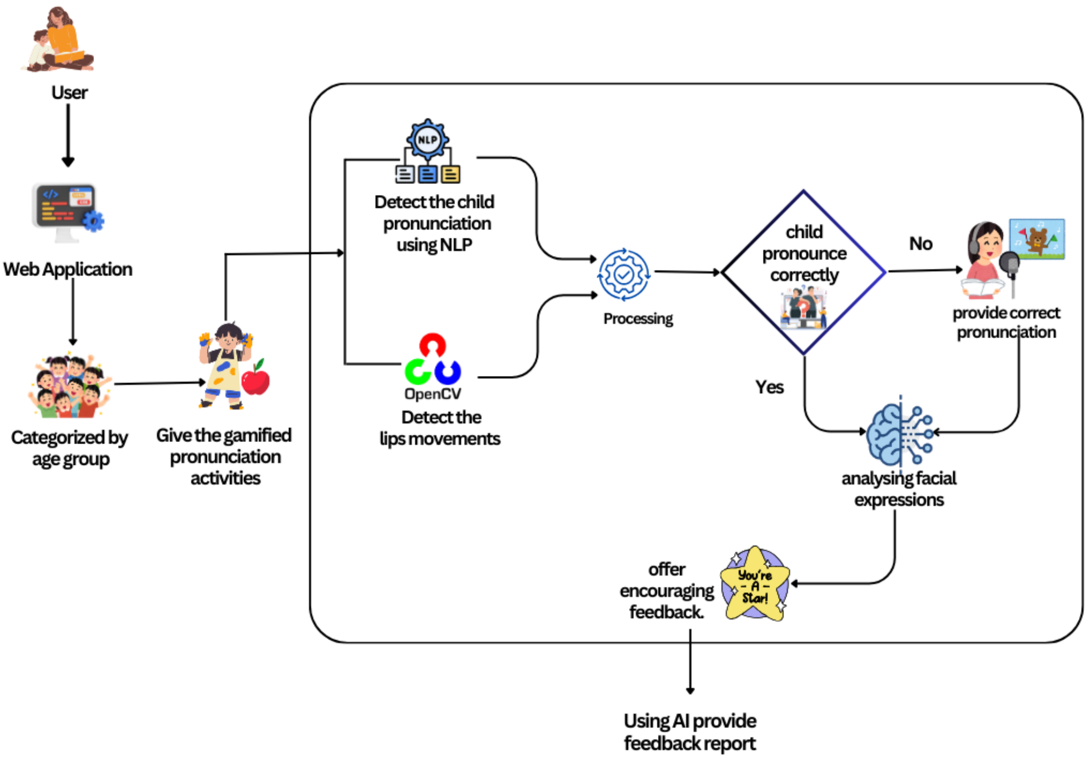

Figure 5 - Component Diagram

The Speech and Language Processing component is designed to support children with Down syndrome, ages 5 to 15, in developing and enhancing their communication skills through a web-based application.

Non-Gamified Component

The non-gamified component is structured to deliver systematic and focused pronunciation training. It encompasses six core lexical categories: Animals, Fruits, Actions, Family, Colors, and Objects. These categories were selected based on their high functional utility in daily communication and their relevance to early language development frameworks. Children are presented with carefully curated visual stimuli paired with corresponding target words, allowing for repeated, distraction-free articulation practice. This mode is particularly effective for learners who benefit from a structured and calm learning environment, emphasizing consistency, visual-verbal association, and progressive reinforcement of correct pronunciation patterns.

Gamified Component

The gamified component introduces an engaging, emotionally responsive learning environment that merges interactive gameplay with real-time pronunciation tasks. A core feature of this module is that the game character advances only when the child exhibits a smile, detected through facial expression recognition, thereby promoting positive effects and emotional engagement. At 10-second intervals, a new image is displayed, prompting the child to articulate the corresponding word. Progression within the game is contingent upon accurate pronunciation, which is validated via integrated speech recognition technology. This design fosters intrinsic motivation, sustained attention, and active participation. Moreover, by requiring smiling for character movement, the game serves a dual purpose: enhancing speech production while supporting emotional regulation. This is particularly beneficial for children who may exhibit elevated levels of frustration, aggression, or social withdrawal, offering a therapeutic channel to encourage positive behavioral responses within a joyful, goal-oriented setting.

Monitoring Facial Expression

By monitoring the facial expressions of the child at the conclusion of every activity, interpret his/her state of emotion (whether happy, sad, or fearful). Following such an interpretation, provide personal feedback such as "You can do it!" or "You are a star!" to ensure positive reinforcement.

Knowledge Test

Test the child's knowledge by asking questions based on the activities done. Ask random questions of the same category to test the extent of understanding. Keep a record of the scores and prepare detailed reports of how they are progressing and areas where improvement is needed.

AI-Driven Feedback Reporting System

The system integrates an AI-powered feedback generation module using Gemini AI to provide personalized progress reports for parents and teachers. Quiz data—such as correct and incorrect answers, time spent, frequently mispronounced words, and engagement patterns—is stored in a secure database. Gemini AI analyzes this data to generate meaningful feedback, including performance summaries, pronunciation challenges, progress trends, and tailored recommendations. In the gamified section, emotional engagement, such as smiling frequency, is also considered to support emotional well-being. This approach ensures that caregivers receive timely, data-driven insights to support each child’s unique learning journey.

2.1.2 Flow chart

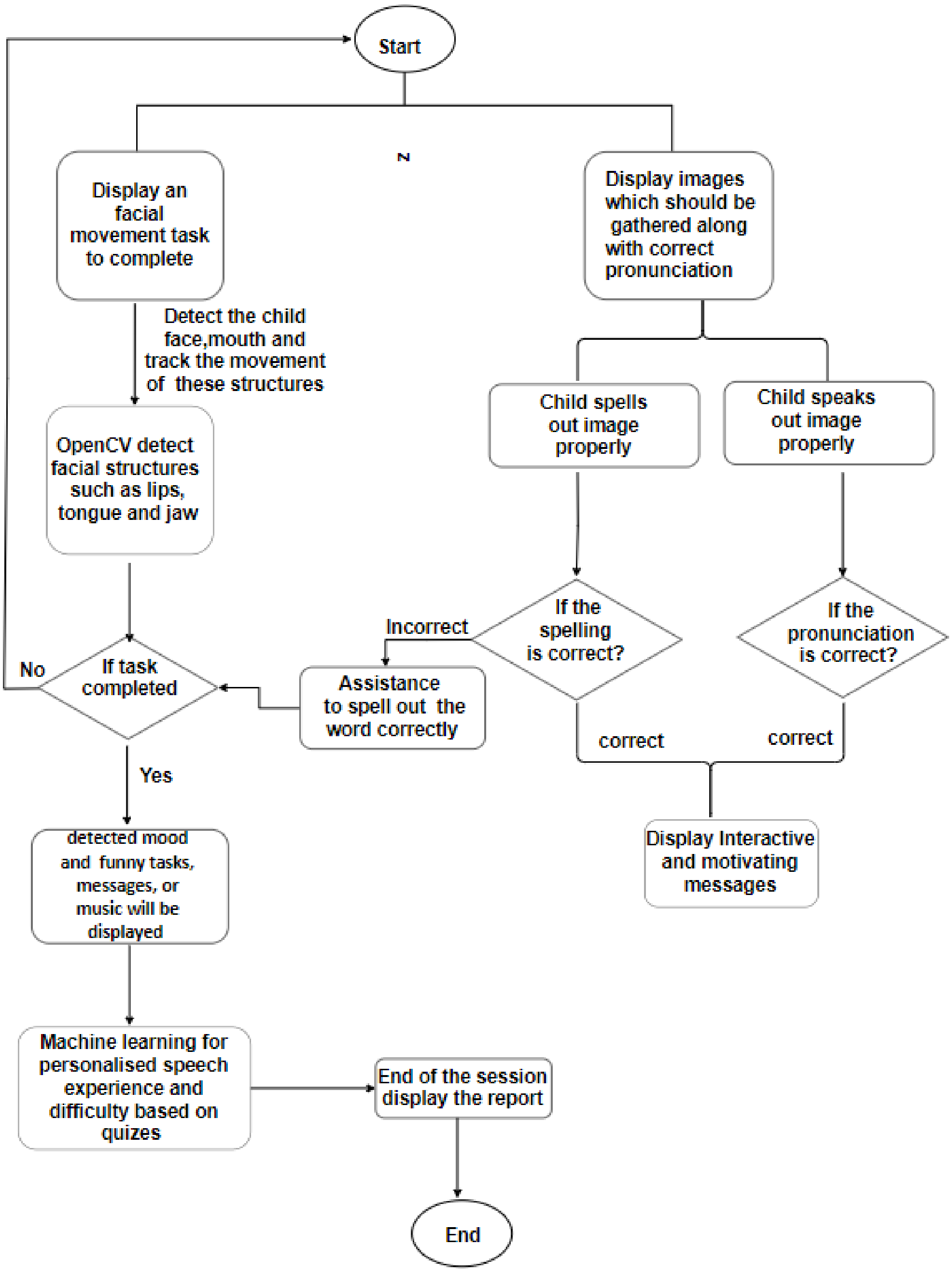

Figure 6 - Flow Chart

2.1.3 Software solution

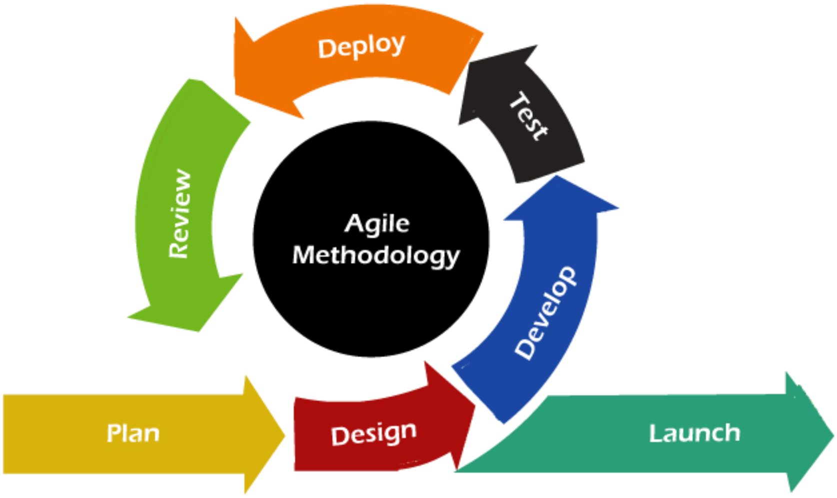

Figure 7- Agile Methodology

Agile methodology is applied as a flexible and iterative way of project management and software development with focuses on continuous improvement and adaptability. Teams plan, design, develop, test, and review features by working in iterative cycles, making them adapt to changing requirements and feedback. Agile places strong emphasis on collaboration and open communication among the team members and stakeholders to resolve any difficulty or realize any opportunity. Due to its emphasis on customer feedback and iterative progress, Agile gives teams the ability to ensure the delivery of a high-quality product that will meet the changing needs of users.

Planning: At this phase, you would describe in detail what objectives are to be attained through activities pertaining to speech and language therapy. This will outline what kind of activity needs to be incorporated for children with Down's syndrome, such as articulation practice or sentence construction, and the technical requirements needed, such as speech recognition and interactive scenarios

Design: Design how the activities will be presented within the web application: it implicates mockups of the interactive elements, layout for exercises in a way to guarantee accessibility and engagement for children with Down syndrome

Development: Develop the functionalities of speech therapy activities, including interactive visual cues for articulation practice, a matching game on daily needs, and advanced sentence construction modules. This shall also involve integrating speech recognition and feedback mechanisms.

Testing: Test speech therapy activities with the target users or simulate their interaction. Check if characteristics such as speech recognition accuracy, interactivity, feedback systems, and others operate according to expectations. Conduct user testing with children to ensure that the activities are interesting and effective.

Deployment: Run the web application with integrated speech therapy activities. Test everything working well in the live environment: users can log in, access, and use the activities without any problems.

Maintenance: Keep a constant check on the application for any problems or user feedback about the speech therapy activities. Update and improve the activities continuously based on feedback from children, parents, and therapists on the effectiveness of the activities and user experience.

2.1.4 Requirements gathering

Conducting Interviews

Interviews of parents, caregivers, and speech therapists will be conducted to increase understanding of specific needs and preferences for speech therapy in children with Down syndrome. Focus groups could reveal perceived barriers against effective communication, current practices, and the efficacy of interventions, alongside existing gaps in available tools. Specific questions about daily challenges in communication, experiences of therapy, and features they would like to see within a web application will help a developer collect rich data to both inform the design and functionality of the application.

Online survey

Surveys provide a broader view of what families and professionals using speech therapy may want and need. Sharing the survey with more people provides quantitative data on common issues, which techniques tend to work well for therapy, and which features are most sought after in a web-based tool. Many of these questions would revolve around the frequency of speech difficulties, types of interactive activities preferred, ratings of effectiveness of existing resources, etc. It will also analyze the results of surveys to help in prioritizing features and tailoring the application to wider community needs.

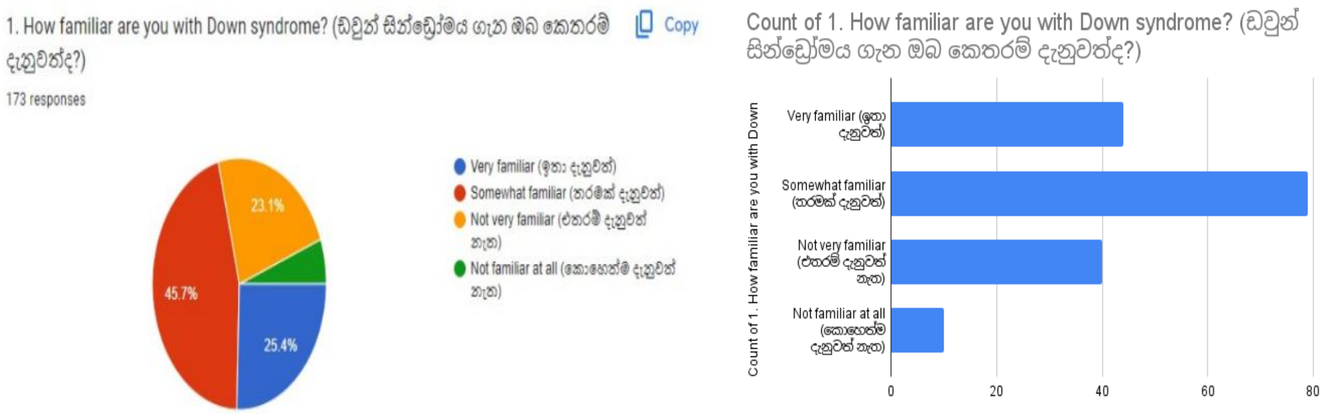

Figure 8 - Survey Question 01

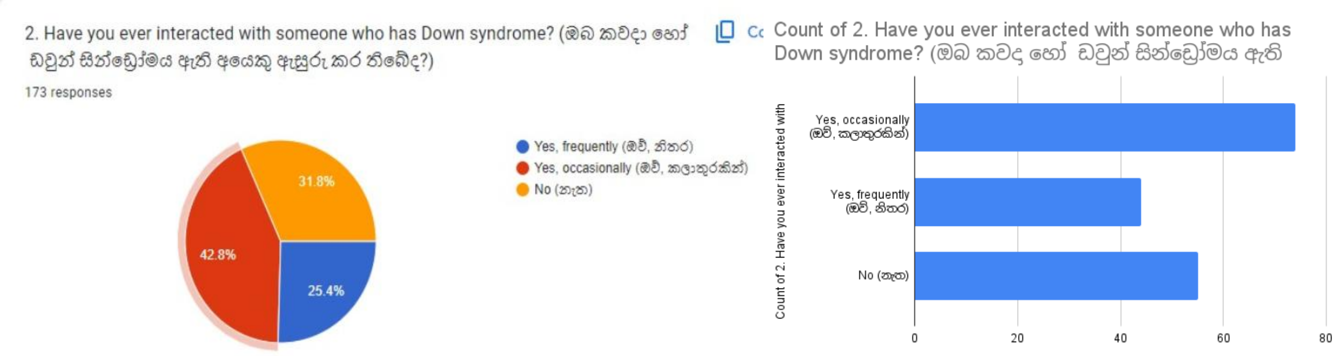

Figure 9 - Survey Question 02

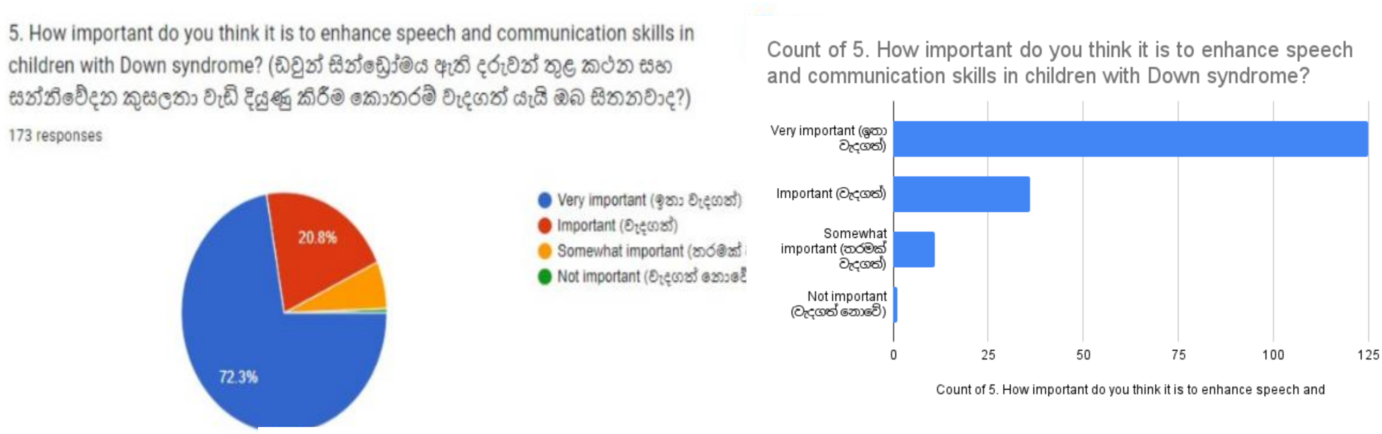

Figure 10 - Survey Question 03

2.1.5 Functional requirements

- **Speech Recognition:** The system shall support the ability of speech recognition for the articulated words to be able to process through Natural Language Processing for the purpose of pronunciation assessment.
- **Real-Time Feedback:** The system must provide feedback at the same time it needs to be output for the pronunciation articulation exercise by using visual and audio indicators that will assist the child to adjust pronunciation.
- **Motion Analysis:** Facial expression and mouth movement are monitored and analyzed in articulation exercise by the ability of image processing.
- **Tracking Progress:** The app is to save the child's data and track his or her progress over time by showing the improved areas and weak ones.
- **Interactive, gamified quizzes**: Make learning and tests interesting through interactivity.
- **Data Storage:** User data is to be stored safely, with the record on progress reports and history of accrued sessions.
- **Accessibility Features:** Making the application user-friendly with people having other needs by incorporating features such as text-to-speech for instructions and, where necessary, visual aids.

2.1.6 Non-functional requirements

- **Performance:** Very minimal latency shall be needed in real-time processing since responses are immediately required.
- **Usability:** It shall be intuitive, easy to use, and helpful across users diversified enough.
- **Scalability:** This technology allows every number of users and massive data without any loss of performance.
- **Security:** Data is secured through data encryption and secure authentication access.
- **Reliability:** The chances of downtime are very thin, with regular maintenance and updates.

2.1.7 Software requirements

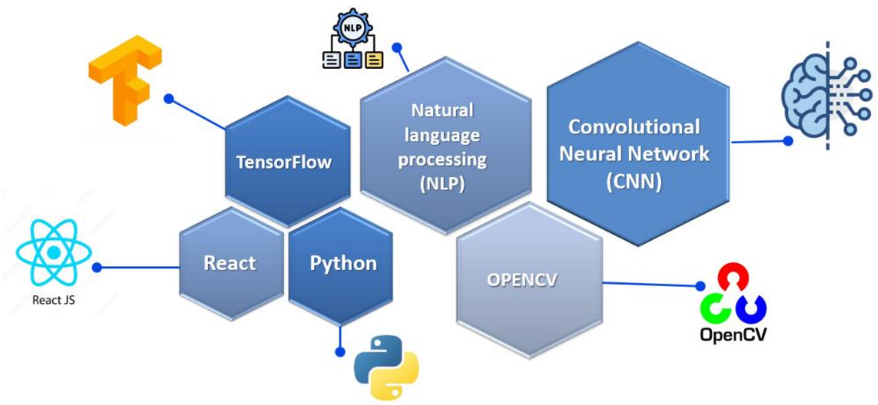

Figure 11- Technologies

Natural Language Processing (NLP): This is the key reason why NLP plays such an important role in the proper recognition and justified categorization of child voice and should supersede other forms of speech analysis, which can be as nonspecific for children with Down syndrome. In contrast, simpler speech recognition systems may face the challenges of multiple pronunciation, which NLP would address well due to the different pronunciation and intonations, thereby giving specifically categorized feedback. This ensures that the application can deal with specific issues of articulation and pronunciation, and NLP is being applied for personalized speech therapy.

OpenCV: This library is used because it processes images well, which is very useful for the analysis of facial expressions and lip movements. In comparison with other image processing libraries that may not provide such detail or real-time performance, OpenCV provides robust tools to capture and analyze visual data. Its versatility in dealing with different image data types makes it perfectly aligned to support the interactive and effective implementation of speech exercises; other less focused image processing tools may not be potent for this one application.

Convolutional Neural Networks (CNNs): CNNs are borrowed for their superior ability in the extraction and analysis of features in visual data, such as lip movements and facial expressions. While many other types of neural networks, or in fact, even simpler machine learning models, may not be as good at dealing with detailed and complicated patterns of visuals, CNNs have been constructed for the purpose of dealing with those hierarchical features in images. This makes them the best choice for not-so-less accurate feedback, particularly in the case of speech-related visual cues.

TensorFlow: TensorFlow is used for its strong machine learning capabilities, which enable adaptive quizzes and building personalized exercises. The flexibility of TensorFlow for supporting models is another reason for its preference. Other machine learning frameworks are either not scalable or flexible, or they might not support a diversity of models and can operate over complex data, but TensorFlow can. Such flexibility in its ecosystem embeds dynamism in building interactive content, in response to the learner's pace; hence, it is more attuned to a framework with enhanced possibilities for personalized learning experiences than less versatile or more basic ones.

Hardware Requirements

Devices: tablets, or computers with cameras and microphones.

Processing Power: At a level that can accomplish real-time speech and image processing

Network Requirements

Internet Connection: for downloading updates, accessing online resources, and storing data in the cloud

2.1.8 User requirements

- **Children with Down Syndrome**: It should be user-friendly and engaging, self-explanatory, and with minimal navigation and interactive elements appropriate to their level of development.
- **Parents/Caregivers**: Progress monitoring, reporting, and setting options for individualized learning.
- **Speech Therapists/Educators**: Designing intervention programs, intervention implementation, tracking progress, feedback.
- **Technical Support**: Help resources and troubleshooting guides to help users in case of a problem.

2.2 Testing & Implementation

2.2.1 Implementation

The system is built using a full-stack architecture that combines frontend interactivity with backend intelligence. The backend logic is implemented in Python using Flask, while the frontend is developed with React.js. The application integrates several technologies, including machine learning (TensorFlow/Keras), natural language processing (phonetics library for pronunciation comparison), MongoDB for data storage, and Google Gemini API for generating intelligent feedback. The core idea is to provide a gamified, engaging pronunciation platform supported by emotion recognition, pronunciation accuracy checking, and weekly performance reporting.

Emotion Detection Model Training

The emotion recognition system is built using Python with the TensorFlow and Keras libraries. A Convolutional Neural Network (CNN) is designed and trained using the FER2013 dataset, which contains labeled facial images for different emotions like happy, sad, angry, etc. The model uses multiple Conv2D layers with ReLU activation, followed by MaxPooling2D, and Dropout layers to prevent overfitting. The Flatten and Dense layers convert image features into a probability distribution using Softmax for multi-class classification. For dataset manipulation, pandas and numpy are used. Model training is optimized using the Adam optimizer and a categorical_crossentropy loss function. The final model is saved using .h5 format and is loaded in the Flask backend for inference.

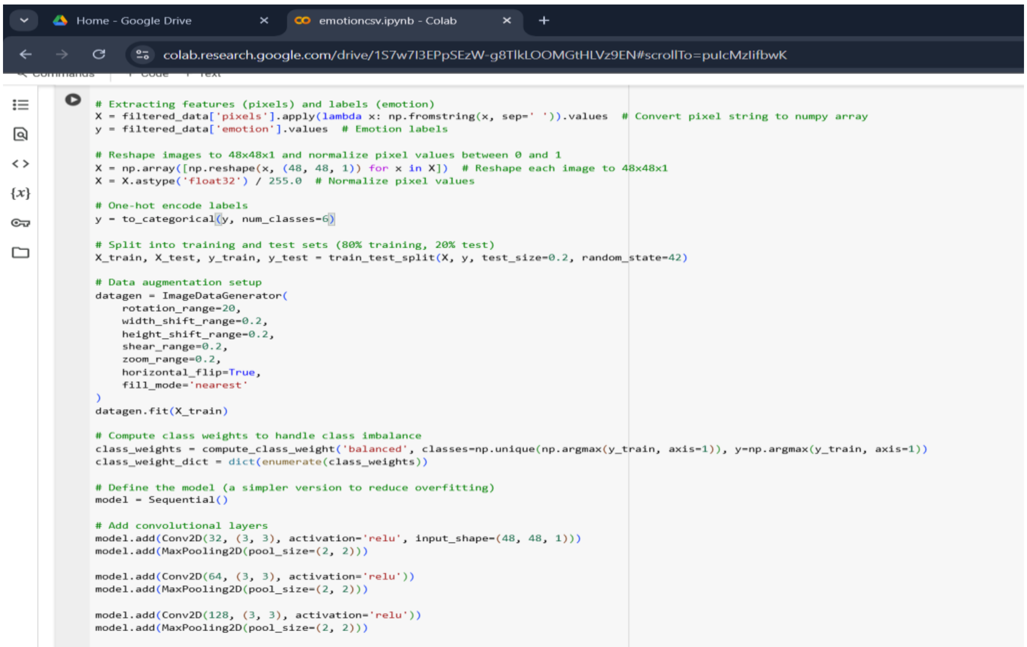

Figure 12- Model training

Voice-to-Text Translation and Pronunciation Checking

The pronunciation checking module uses JavaScript (React) on the client side and Python (Flask) on the server side. The browser's native SpeechRecognition API (webkitSpeechRecognition) captures audio input and converts it to text. This text is sent to a Flask backend using **Axios** HTTP POST requests. On the server side, the phonetics Python library is used to compare the pronunciation of the user's word with the target word using the Metaphone algorithm, which converts both words into phonetic encodings to evaluate similarity. If the encodings match, the pronunciation is marked correct; otherwise, feedback is returned with suggestions.

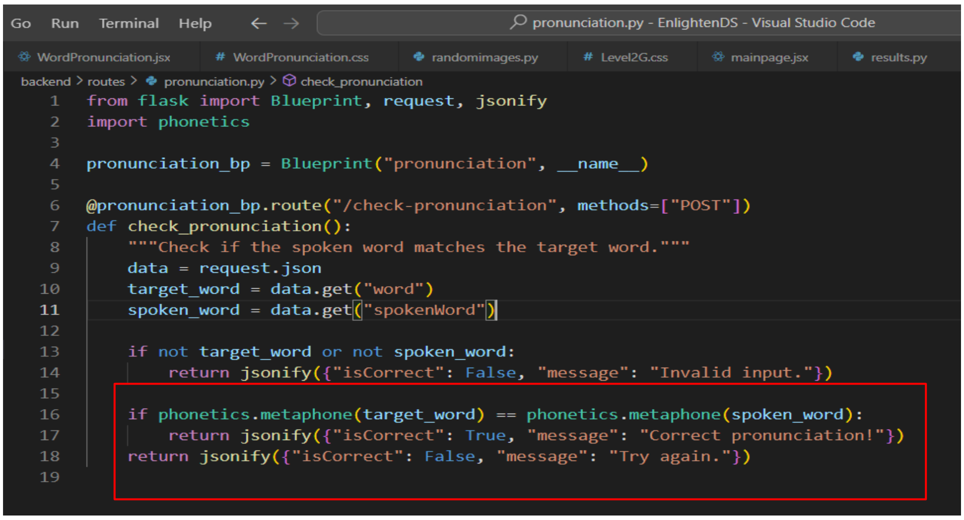

Figure 13 - Pronunciation-backend route

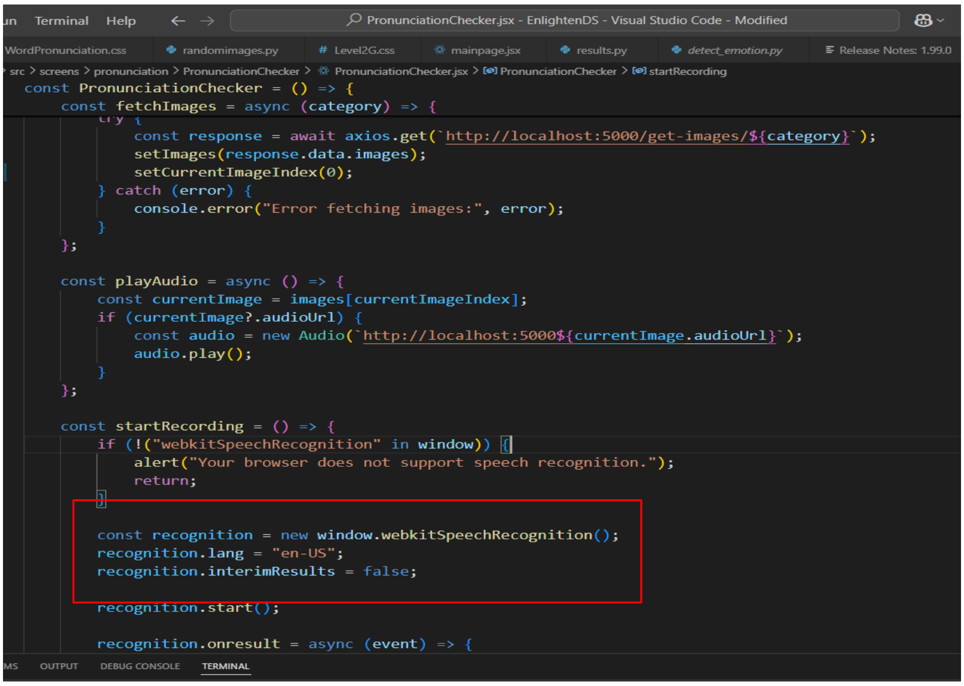

Figure 14 -Pronunciation frontend route

User Interface

The frontend is implemented using React.js, providing a highly responsive, single-page application. Components like Quiz, Webcam Capture, and Feedback Modal are built using functional components and useState, useEffect hooks. For voice input, the Web Speech API (webkitSpeechRecognition) is used to capture real-time pronunciation attempts. For facial emotion analysis, we use the **react-webcam** library to capture webcam frames. These frames are sent to the Flask backend at intervals using setInterval + Axios, where they are processed for emotion classification. Based on the result (e.g., smiling face), motivational feedback like “Well done!” or “Keep going!” is displayed.

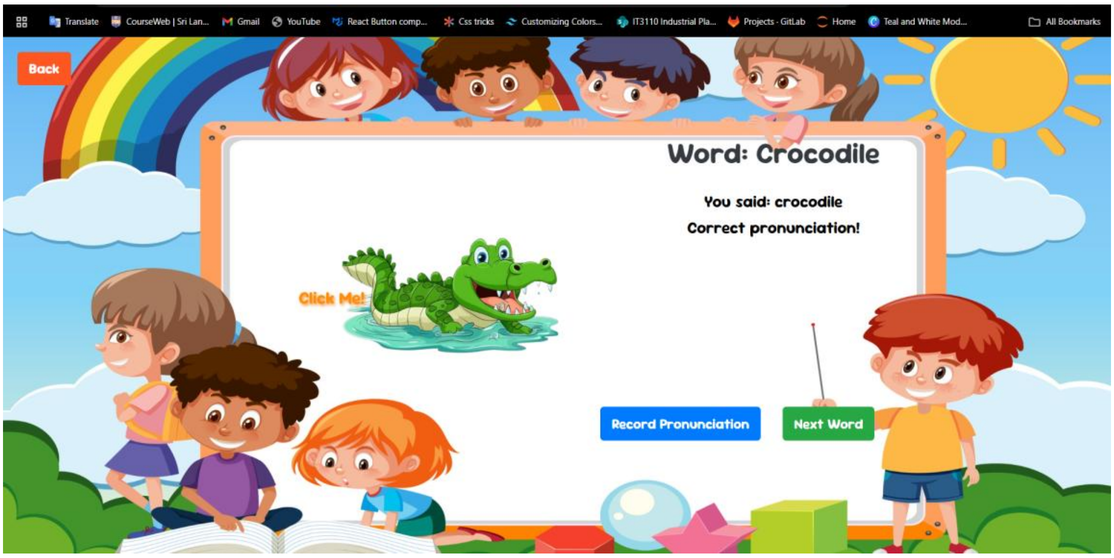

Figure 15 – User Interface 01

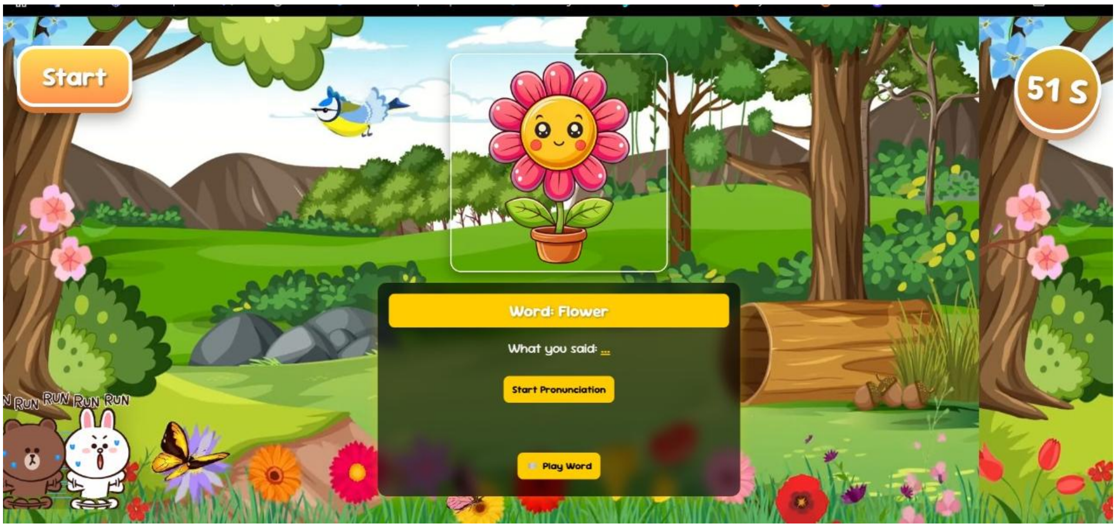

Figure 16- User Interface 02

Figure 17- User Interface 03

Database Integration for Progress Tracking

To monitor user performance and enable feedback generation, MongoDB was integrated using the pymongo library. A secure connection was established using environment variables stored in a .env file. The results collection within the quiz_app database stores structured quiz records, including scores, timestamps, categories, and user identifiers. This persistent storage is critical for generating long-term insights into each child's progress.

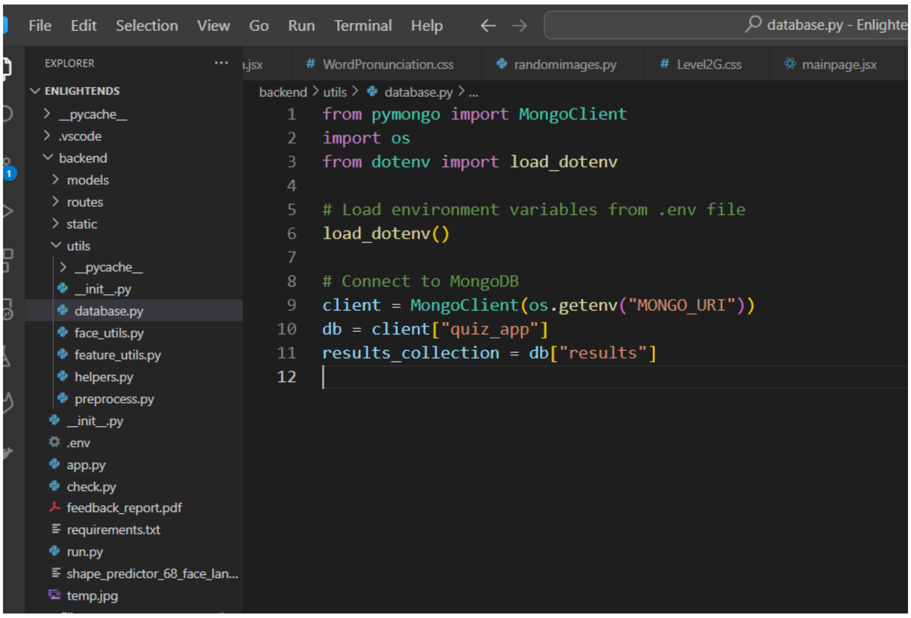

Figure 18 - MongoDB Connection

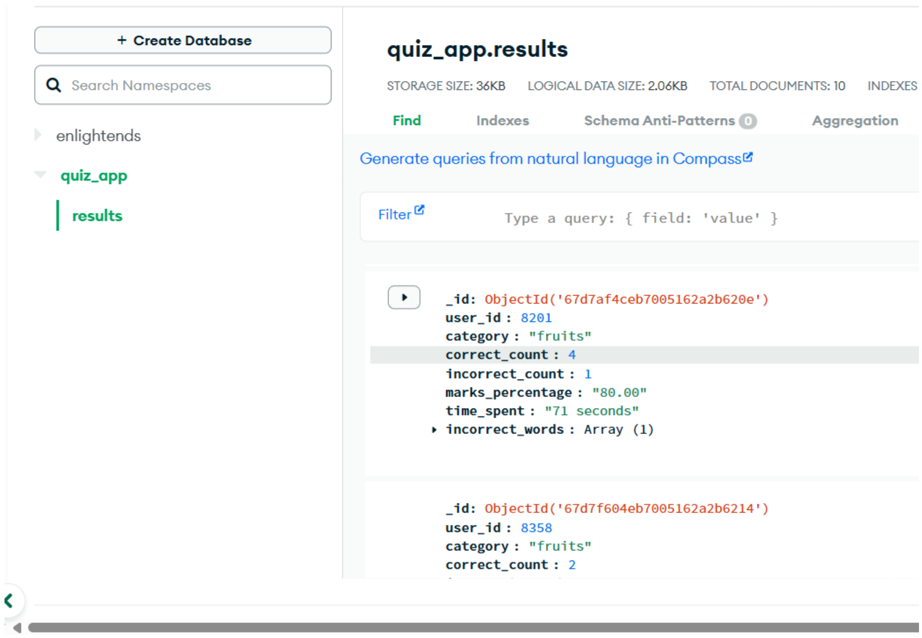

Figure 19 - Data Collection

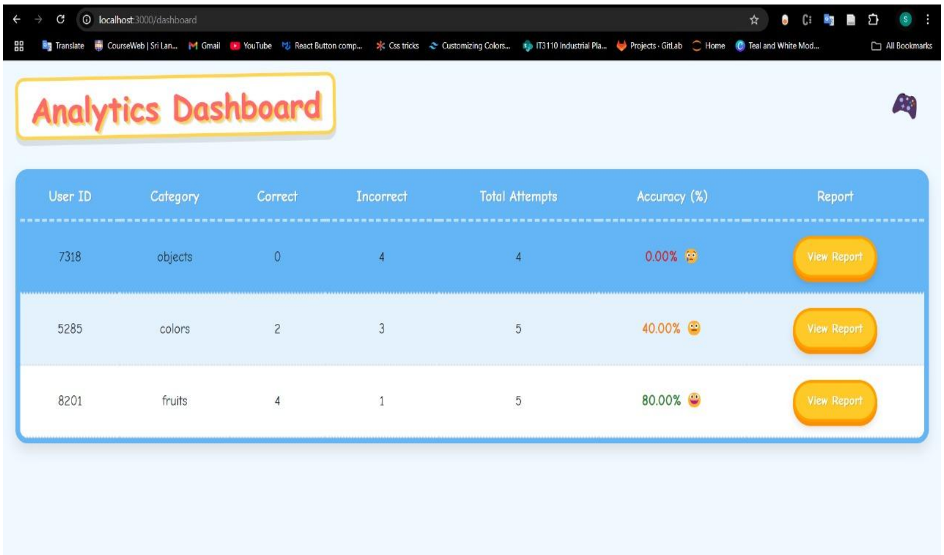

Figure 20 - User Interface 04

Personalized Report Generation using Gemini API

To generate personalized feedback after each session, we use the Google Gemini API. The backend script queries quiz data from MongoDB, aggregates information (e.g., frequent mistakes, average score, time spent) and constructs a prompt. This prompt is sent via a POST request to the Gemini API endpoint using Python's requests library. The LLM responds with child-friendly and parent-understandable feedback, which is stored in MongoDB and displayed on the frontend. The goal is to provide a report that highlights strengths, identifies weaknesses (e.g., struggles with "th" sounds), and recommends simple at-home exercises.

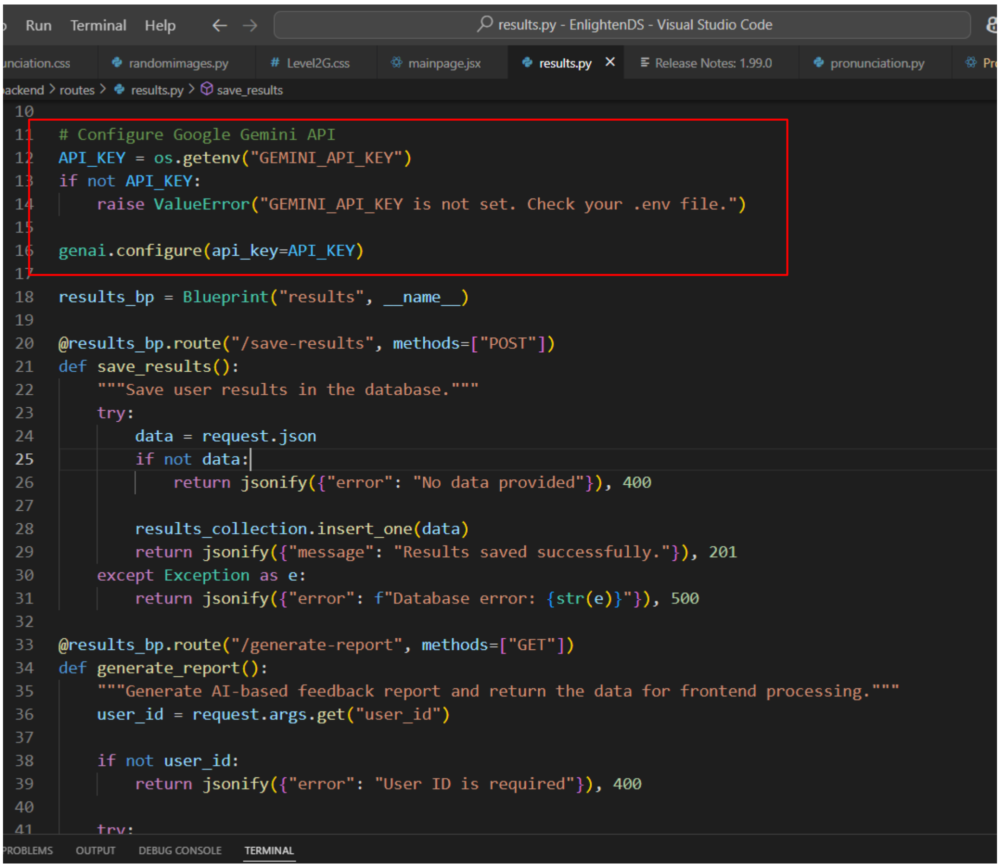

Figure 21- AI Feedback backend

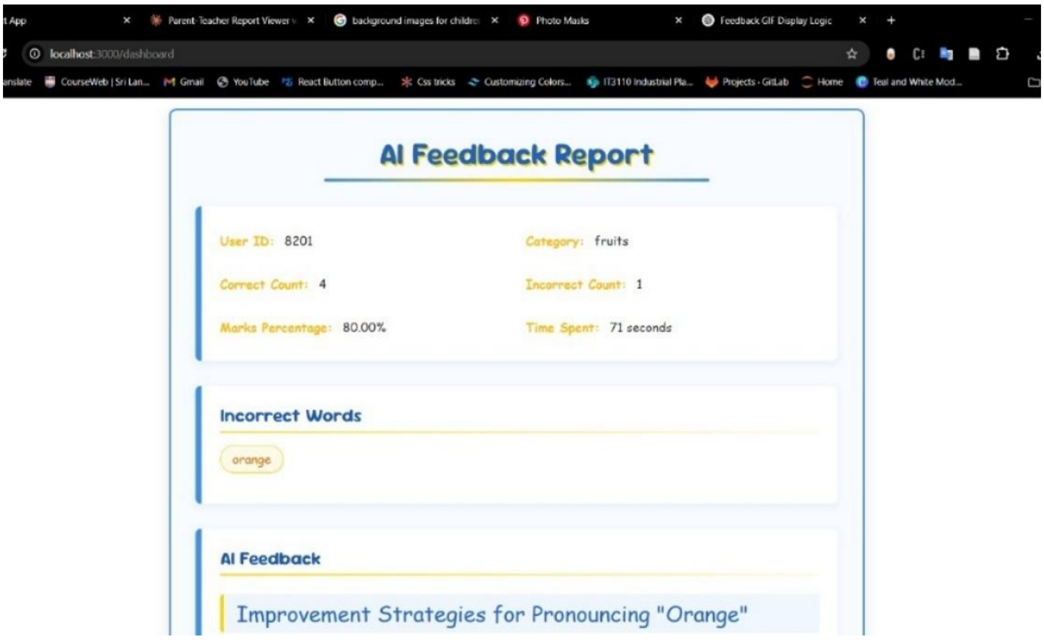

Figure 22 - Feedback Report UI

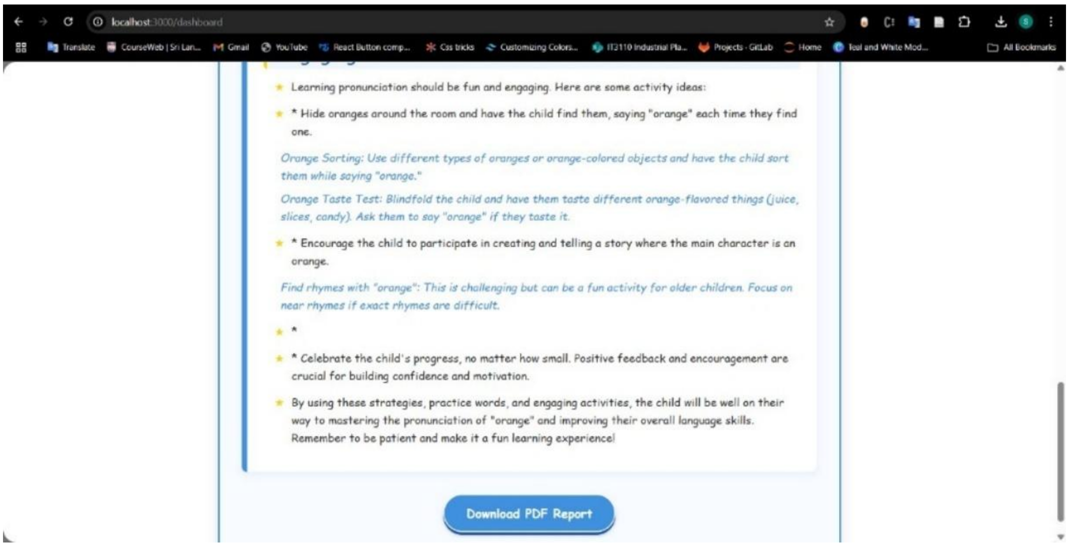

Figure 23- Feedback Report

Gamified Environment with Emotion Feedback

To improve engagement, the app includes a gamification layer driven by emotion recognition. The emotion detection model (served via Flask) receives webcam images from the frontend and classifies the user's emotion. If the detected emotion is positive (e.g., "Happy" or "Excited"), the system dynamically increases a motivational score, displays visual animations (stars, emojis), or plays audio rewards using JavaScript Audio API. This emotional intelligence integration is designed to keep children motivated and reduce frustration during speech learning.

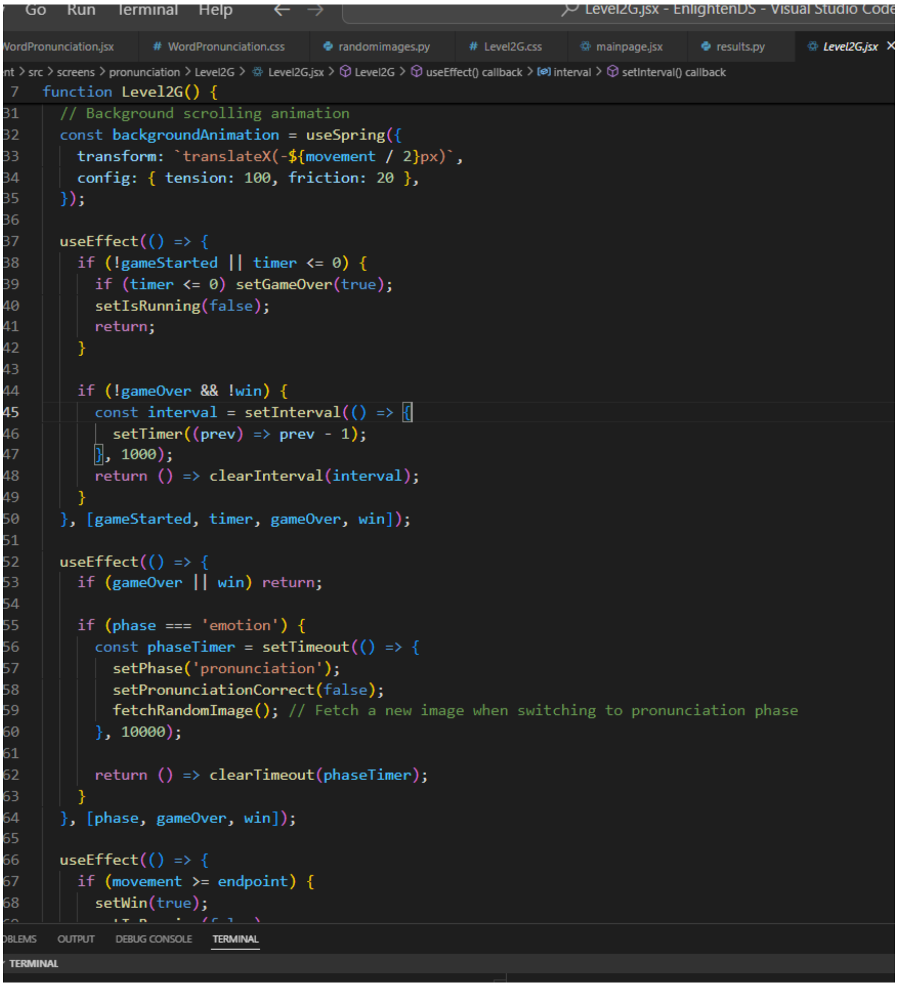

Figure 24 - Gamified Quiz

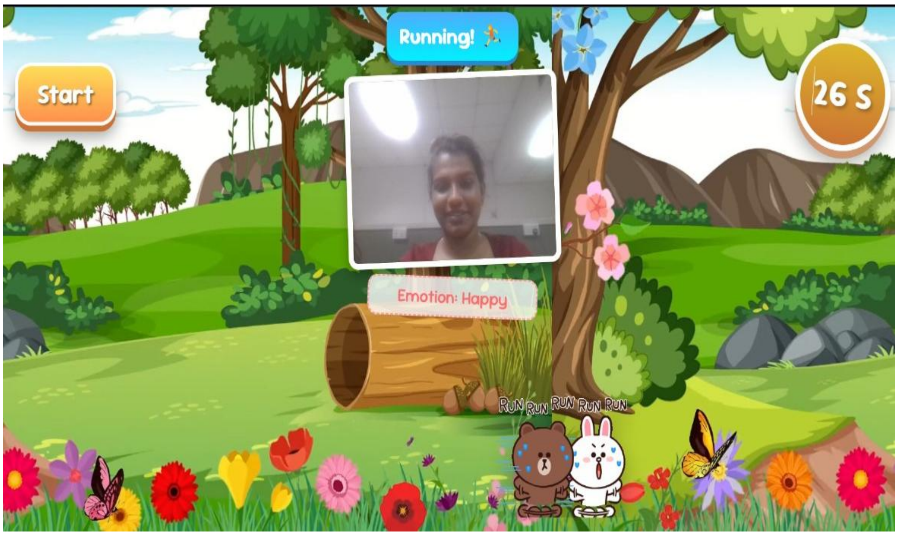

Figure 25- Gamified UI

2.2.2 Development tools

Visual Studio Code (VS Code) – Code Editor & IDE

Visual Studio Code (VS Code) is the primary Integrated Development Environment (IDE) used during the development of the application. It supports both frontend (React.js) and backend (Flask/Python) development in a single workspace. VS Code offers robust features like IntelliSense for code completion, syntax highlighting, built-in terminal, Git integration, and debugging tools. Extensions such as Python, Prettier, ESLint, React Snippets, and MongoDB for VS Code were used to speed up development and maintain consistent code quality. The modular file structure — with folders like frontend, backend, models, and routes — was organized using VS Code's Explorer, allowing seamless project navigation and management.

GitHub – Version Control & Collaboration

GitHub is used for version control, team collaboration, and source code management. The project was initialized with a Git repository using the git init command and then pushed to a GitHub repository. Branching strategies such as main, dev, and feature/emotion-model branches were used to manage parallel development workflows. GitHub Actions was also considered for continuous integration (CI) to automate testing or deployment in the future. Pull requests allowed team members to review code and merge changes collaboratively. Issues and projects features were used to track bugs, tasks, and milestones, ensuring Agile development practices.

Google Colab – Model Training & Experimentation

Google Colab was utilized primarily for training the emotion detection CNN model and experimenting with data preprocessing, image augmentation, and model evaluation. Colab provides a cloud-based Jupyter notebook environment with free access to NVIDIA GPUs, significantly reducing training time. The tensorflow, keras, and opencv-python libraries were installed within the notebook using pip. Trained models were saved in .h5 format and later exported to the Flask backend for real-time inference. Colab notebooks were versioned and stored on Google Drive, making it easier to share and reproduce experiments among team members.

2.2.3 Testing

- Pronunciation Checker API Testing

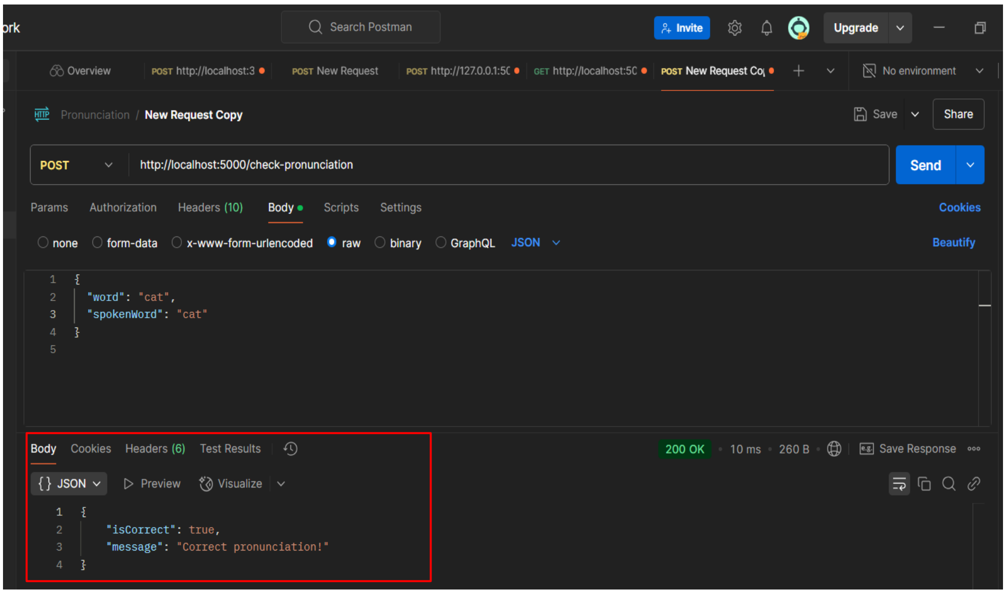

Figure 26 - Testing_01

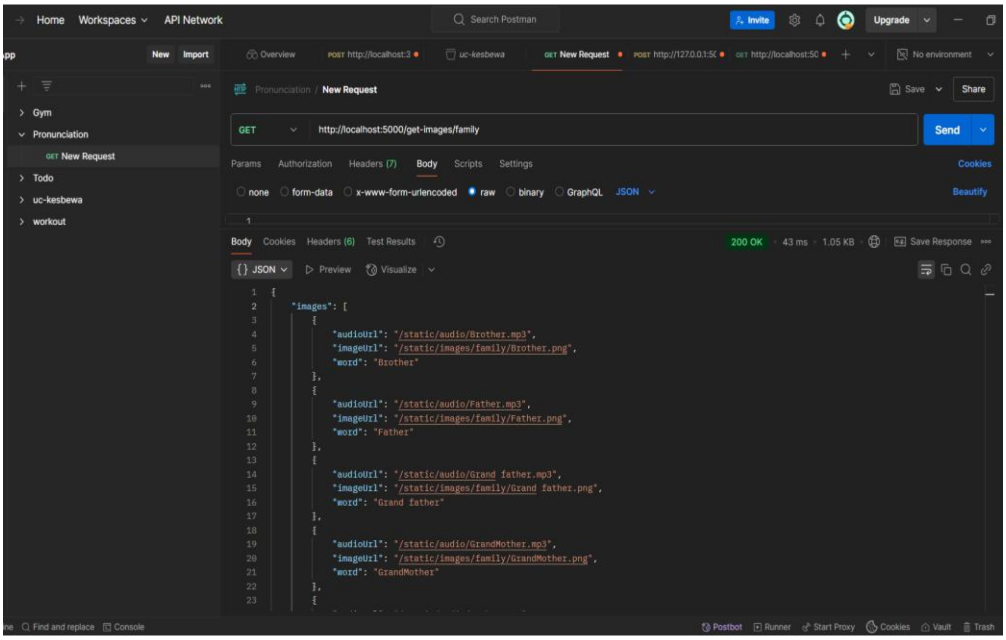

Figure 27- Testing_02

- Emotion Detection API Testing

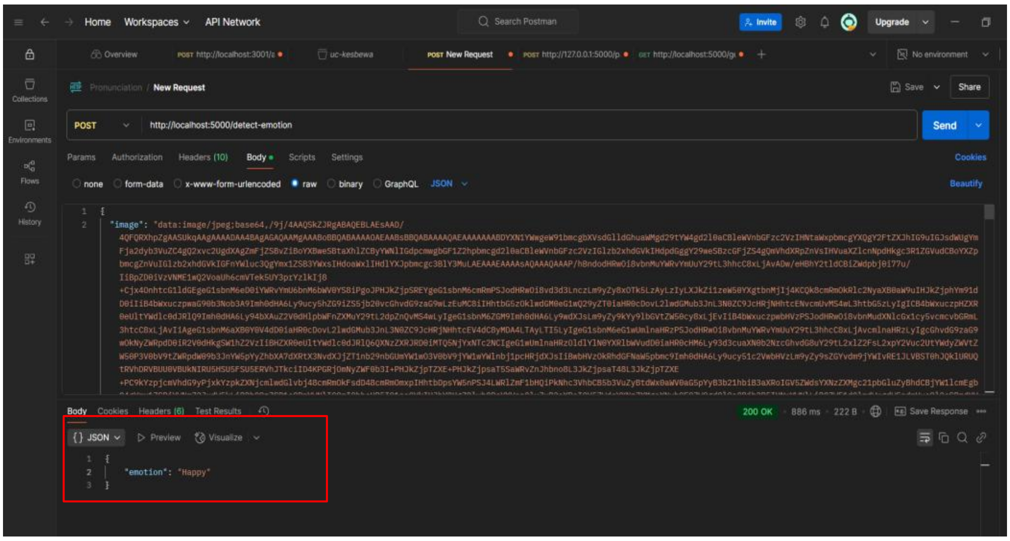

Figure 28- Testing_03

- User Data Storage API Testing

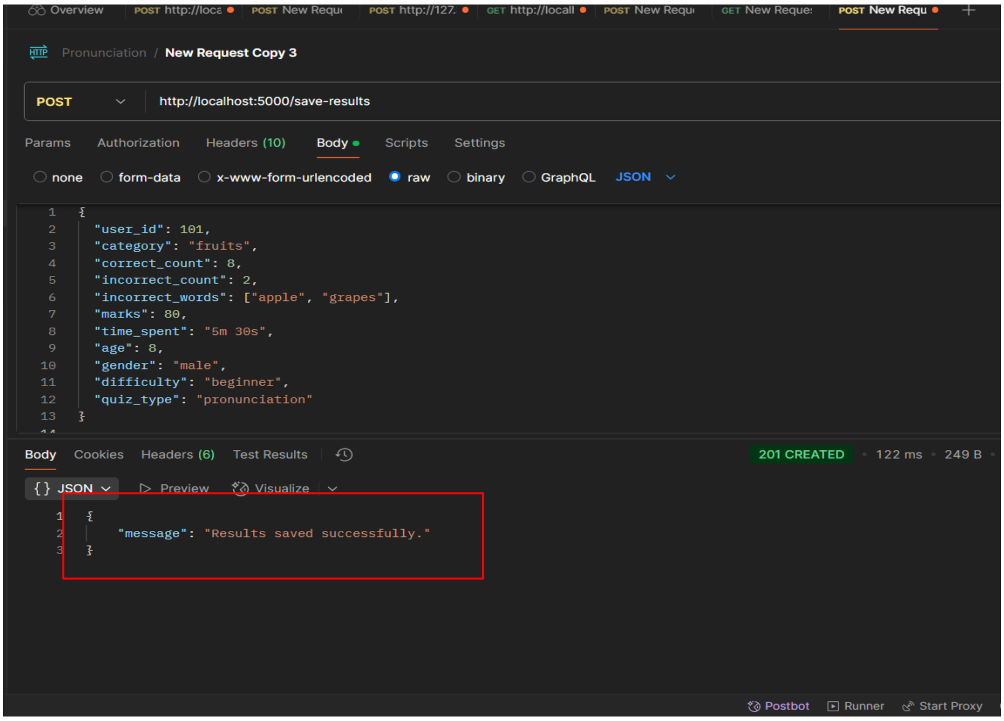

Figure 29 - Testing_04

- Feedback Generation API Testing

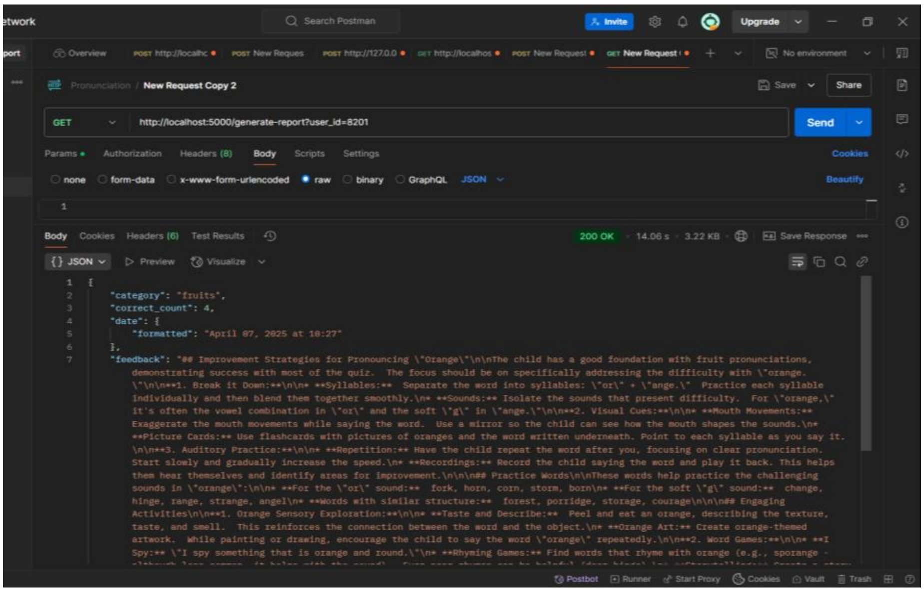

Figure 30- Testing_05

2.3 Commercialization Plan

Market Analysis

Target Audience

- Children with down syndrome age 5-15
- Parents, caregivers and schoolteachers
- Health care professionals working with children with special needs

Market Size and Trends

- Increasing awareness and demand for individualized educational help for children with special needs and increasing adoption of technology in the learning process, especially in special education

Revenue Model

Freemium Model

- Offer a free, basic version of the system, with reduced features.

Subscription-Based Model

- Users could have a subscription service on a monthly or annual basis for parents and institutions and have tiered pricing based on features/levels of access (basic, premium)

Institutional Sales

- Partner with schools and educational institutions by offering bulk subscriptions.

Packages and features

Basic Plan

- Features
  - ✓ Access to cognitive level assessment tools
  - ✓ Basic communication skill improvement activities
  - ✓ Limited personalized quizzes for math skills
  - ✓ Access to painting and piano platforms to develop creativity
  - ✓ Basic summary reports for parents with scores and spent time
- Price: Free

Premium Plan

- Features
  - ✓ All features from the Basic Plan
  - ✓ Advanced cognitive assessment with detailed symptom analysis
  - ✓ Full access to personalized quizzes with adaptive difficulty
  - ✓ Emotion-based engagement activities and mood detection
  - ✓ Advanced summary reports with performance improvement statistics.
  - ✓ Priority customer support.
- Price: $10 per month

School/Institution Plan

- Features
  - ✓ Bulk subscriptions with discounts for schools and educational institutions.
  - ✓ Access up to 40 students.
  - ✓ Full access to all features for multiple users [Students].
  - ✓ Training and support for educators to include it in the curriculum.
  - ✓ Detailed analytics and reports for educators on student progress
- Price: $30 per month

2.3.1 Budget

Component	Amount (LKR)
Travelling cost	10000
Server and hosting charges	25000
Internet charges	15000
Total	50000

Table 2 - Budget

The budget for "EnlightenDS" has been carefully planned to cover all the important costs needed to develop and launch our cognitive assessment tool. It involves an amount of LKR 50,000, carefully planned to cover all the essential costs associated with the development and deployment of our cognitive assessment tool, EnlightenDS. The total budget is broken down into three key components.

First, we allocated LKR 10,000 for Traveling Costs. This amount is for traveling related to team meetings, collaboration with educational institutes, and any fieldwork that might be required for collecting user feedback or on-site demonstrations of our tool. This will ensure that we get to meet and interact face-to-face with our stakeholders. This is very important in the refinement of our solution to meet real-world needs.

This comes with the largest portion of the budget, LKR 25,000, for Server and Hosting Charges. This will be very instrumental in maintaining the infrastructure required to support our app. Reliable server hosting ensures that our platform remains open to users always, ensuring seamless operation and data security.

Finally, LKR 15,000 has been allocated for Internet Charges. Since we are going to use cloud-based services and collaborate remotely, stable internet with high speed is one of the essential requirements for the day-to-day running of the project from research to development, testing, and communication.

2.3.2 Gantt chart

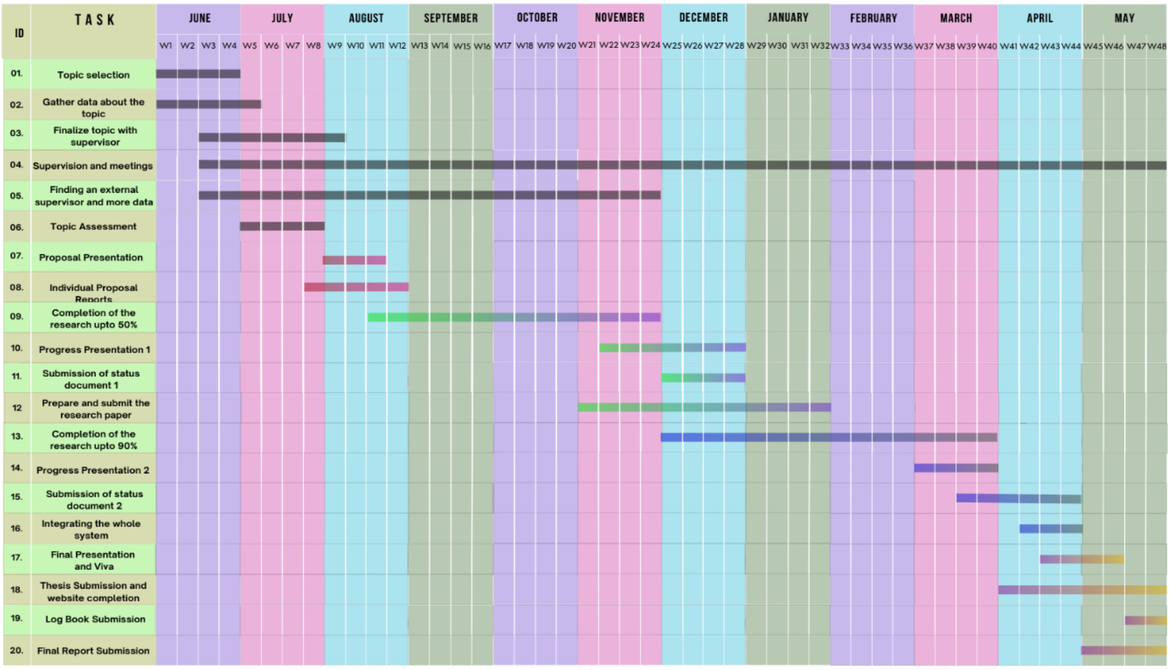

Figure 31 - Gantt Chart

2.3.3 Work breakdown chart

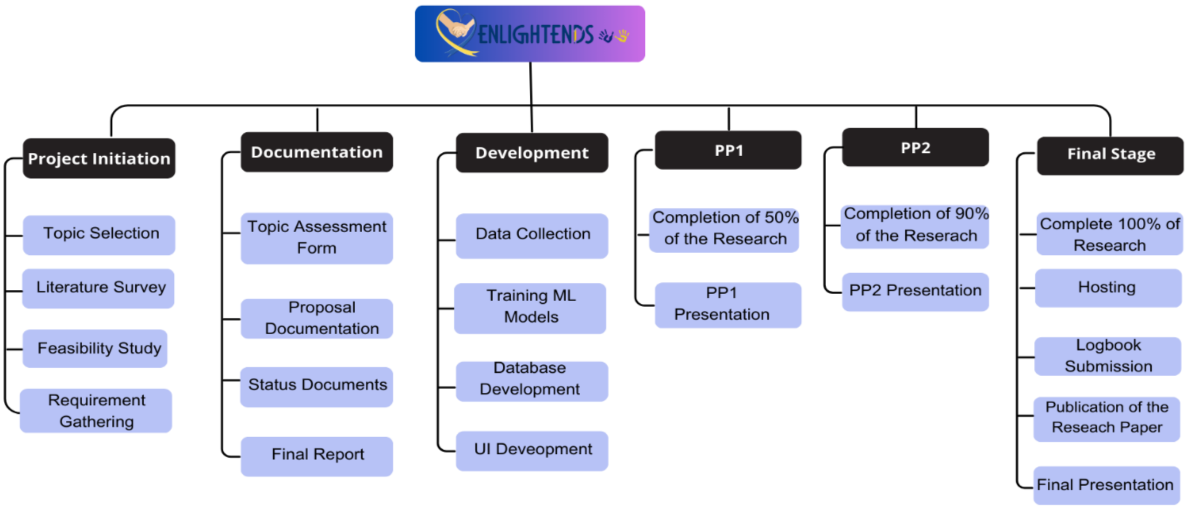

Figure 32 - Work breakdown chart

3 RESULTS AND DISCUSSION

3.1 Results

The proposed pronunciation enhancement system demonstrated significant improvements across several key areas, including phonetic accuracy, vocabulary retention, engagement levels, and learning progress. The integration of advanced technologies such as computer vision, emotion recognition, real-time tracking, and interactive learning tools yielded promising results. The following outcomes were observed:

- **Emotion Recognition**: The emotion recognition system, utilizing Convolutional Neural Networks (CNNs) and OpenCV, achieved a classification accuracy of **85%**. This allowed the system to detect the emotional state of the children, enabling dynamic adjustments to the difficulty and pace of learning tasks based on their facial expressions. This adaptive feedback proved essential in maintaining motivation and engagement.
- **Phonetic Accuracy Improvement**: The combination of Google Speech-to-Text and Mozilla DeepSpeech resulted in a 20% improvement in phonetic accuracy among children. Over the testing period, children showed notable progress in correctly pronouncing target phonemes, as the system provided targeted, real-time corrective feedback.
- **Instant Feedback Effectiveness**: The Web Speech API enabled the system to provide immediate feedback on pronunciation, leading to a 15% increase in pronunciation accuracy. By offering real-time corrections, children were able to adjust their speech promptly, reinforcing proper pronunciation patterns and accelerating learning.
- **Engagement Duration**: Emotion-aware adjustments, driven by facial emotion recognition, contributed to a 30% longer interaction time. The system dynamically adapted the difficulty and content based on the child's emotional responses, preventing frustration and enhancing engagement throughout the learning sessions.
- **Vocabulary Retention**: The interactive nature of the system resulted in a 25% increase in vocabulary retention. Repeated exposure to commonly used words, such as those from categories like animals, fruits, and actions, helped children retain the words learned, ensuring long-term recognition and recall.
- **Quiz-Based Learning Enhancement**: The introduction of quizzes, coupled with data tracking and analysis, provided valuable insights into children's progress. The systematic collection of quiz scores revealed consistent improvement trend**s**, helping identify areas of strength and weakness in individual learning curves.

These results underscore the effectiveness of integrating emotion-based learning strategies, multimodal feedback, and personalized tracking to support children with Down syndrome in improving their pronunciation, retaining vocabulary, and maintaining engagement during learning activities.

3.2 Research Findings

The evaluation of the pronunciation enhancement system revealed several significant outcomes related to pronunciation improvement, vocabulary retention, and learner engagement. The findings were derived from testing sessions conducted over a four-week period with children with Down syndrome, aged 5-15 years. Below are the key research outcomes

Pronunciation Improvement

The system showed noticeable improvement in phoneme accuracy. This was assessed by tracking the accuracy of phoneme recognition during testing sessions. Over time, as children regularly interacted with the system, the accuracy improved significantly. This progress can be attributed to the consistent and targeted feedback the system provided on individual phonemes, which helped children refine their pronunciation skills.

Vocabulary Retention

The system demonstrated substantial gains in vocabulary retention. This was evaluated by monitoring the consistency of correct responses during repeated practice sessions. The children were tested on commonly used words in categories such as fruits, animals, and actions. The engaging nature of the system and the repeated exposure to vocabulary helped children retain the words they were learning, ensuring long-term recognition and recall.

Feedback Effectiveness

The real-time feedback mechanism, powered by the Web Speech API, played a critical role in improving pronunciation accuracy. By providing immediate corrective feedback when a child mispronounces a word or phoneme, the system enabled faster learning. This mechanism reinforced correct pronunciation patterns, ensuring that children were continually refining their speech during interactions.

Engagement Duration

A significant increase in interaction time was observed due to the system's emotion-aware feedback. By recognizing facial expressions and adapting to the child’s emotional state, the system dynamically adjusted the difficulty and pace of learning. If the system detected signs of frustration or disengagement, it modified the content to maintain the child’s interest. This led to prolonged engagement, fostering sustained interaction and learning.

Learning Progress

The system demonstrated steady progress in quiz performance, as shown by the data stored in MongoDB. Over the testing period, the children's quiz results consistently improved, indicating the system's effectiveness in supporting ongoing learning. The quiz scores helped track both correct and incorrect responses, providing valuable insights into the areas where children needed more practice. This tracking enabled personalized learning and highlighted the system's capacity to adapt to individual learning needs.

3.3 Discussion

The success of the pronunciation enhancement system can be attributed to the seamless integration of advanced speech recognition, emotion-aware feedback, and interactive learning tools. Google Speech-to-Text and Mozilla DeepSpeech enabled detailed analysis of phoneme accuracy, which was essential for the system's ability to provide precise and effective pronunciation feedback. This, combined with the real-time feedback mechanism from the Web Speech API, ensured that children could hear and correct their pronunciation immediately, facilitating a more efficient learning process.

A particularly innovative aspect of the system was the emotion recognition component, which utilized OpenCV and CNNs to monitor children's facial expressions and adjust the difficulty of the learning tasks accordingly. This emotional adaptability was especially beneficial for children with Down syndrome, as it helped reduce frustration and maintain engagement. By detecting signs of stress, boredom, or frustration, the system dynamically adapted the learning pace and content, leading to a 30% increase in the average duration of learning sessions.

Moreover, the integration of daily quizzes and the tracking of quiz results provided valuable insights into individual learning progress. By storing this data in MongoDB, the system was able to offer personalized feedback, track long-term progress, and generate detailed reports for parents. These reports not only highlighted areas of improvement but also helped identify persistent difficulties, allowing for targeted interventions at home.

However, there are some limitations that need to be addressed to further enhance the system's effectiveness:

- **Environmental Noise**: Background noise in the learning environment posed challenges for accurate speech recognition. To mitigate this, incorporating advanced noise reduction algorithms would enhance speech-to-text conversion, particularly in noisy environments.
- **Facial Recognition Challenges**: The accuracy of emotion detection was occasionally compromised by inconsistent lighting conditions or facial occlusions. Improved algorithms for facial expression detection in varied environments would increase the reliability of emotion-based adjustments.
- **Rare Word Recognition**: Words that were not commonly used in daily conversations presented challenges for both pronunciation and recognition accuracy. Expanding the phoneme dataset to include a wider range of child-specific and disorder-specific speech data would help the system better accommodate these rare pronunciations.

Proposed Enhancements

To build on the current successes, several enhancements are proposed:

- **Noise Reduction Filters**: Incorporating more advanced noise filtering techniques would help improve speech recognition accuracy, particularly in environments with significant background noise.
- **Expanded Phoneme Dataset**: Increasing the diversity of the phoneme dataset to include a broader range of child-specific phonetic patterns would enhance the system's ability to recognize and adapt to varied speech traits, especially those found in children with Down syndrome.
- **Reinforcement Learning**: By introducing reinforcement learning, the system could adapt its feedback and difficulty levels based on real-time user progress. This would enable more personalized and effective learning experiences tailored to each child's unique needs.
- **Augmented Reality Activities**: The addition of AR-based exercises would significantly enhance the learning experience by providing more immersive and interactive opportunities for children to practice vocabulary in context-rich environments. This would not only boost engagement but also provide a more dynamic approach to language acquisition.

These proposed improvements would further refine the system, making it even more effective in supporting the speech development of children with Down syndrome.

Table 3- Test Case_01

Test case ID: Test_01
Test title: Speech Recognition Accuracy Test
Test priority (High/Medium/Low): High
Module name: Speech Recognition Module
Description: This test case is designed to verify the accuracy of the system's ability to recognize and score words spoken by the child.

Pre-conditions: The system has registered the child, and the speech recognition module is functional.				
Test ID	**Test Steps**	**Expected Output**	**Actual Output**	**Result (Pass/Fail)**
Test_01	1. The child speaks a word. 2. Then the system processes it through the speech recognition module.	• System identifies the word correctly and scores it.	• The system identified the word correctly and scored it accurately.	Pass

Table 4 - Test Case_02

Test case ID: Test_04				
Test title: Progress Tracking Test				
Test priority (High/Medium/Low): Medium				
Module name: Progress Tracking Module				
Description: This test case evaluates the accuracy of the progress tracking module in displaying the correct progress reports and related data.				
Pre-conditions: The system has registered the child, and progress data from previous sessions is available.				
Test ID	**Test Steps**	**Expected Output**	**Actual Output**	**Result (Pass/Fail)**
Test_04	1.The system generates a progress report based on the child's activities.	• Progress report is showing correctly; related data is being shown.	• Progress report matched the actual data of exercises performed.	Pass

4 CONCLUSION

In conclusion, this study has successfully fulfilled its primary objective of enhancing pronunciation skills in children with Down syndrome through the development of a highly intelligent and emotionally adaptive learning platform. The platform's integration of state-of-the-art technologies such as speech recognition, real-time corrective feedback, emotion detection, and gamified quizzes resulted in significant improvements in phonetic accuracy, vocabulary retention, and overall user engagement. These advancements not only addressed the immediate speech challenges faced by children with Down syndrome but also created a more engaging and supportive environment for their learning.

A key finding of this research is the crucial role of emotion-aware learning environments in enhancing the effectiveness of speech and language interventions. The platform's ability to assess and respond to the emotional state of the child allowed for tailored learning experiences that were more motivating and engaging. This feature proved to be especially important in maintaining the child's interest and focus, as emotional fluctuations often impact learning progress. By personalizing the learning experience based on emotional and performance data, the system not only promoted more consistent practice but also fostered a deeper connection between the child, the content, and their caregivers. Caregivers were empowered with insights into their child's emotional state and progress, enabling them to adjust learning strategies and provide targeted support in areas where the child was struggling.

The inclusion of real-time corrective feedback was another critical component. Children received immediate guidance on their pronunciation attempts, facilitating faster error correction and reinforcing positive speech patterns. This immediate feedback loop, combined with the system's personalized recommendations for practice, significantly improved both the accuracy of pronunciation and the child's overall confidence in speaking. Moreover, the platform's gamified quizzes contributed to higher levels of engagement, making learning enjoyable and reinforcing key vocabulary through interactive play.

One of the study's most important contributions lies in the seamless integration of multiple technologies, demonstrating that multi-modal interventions can address both speech and cognitive challenges faced by children with Down syndrome. This approach has the potential to serve as a model for future language-learning platforms aimed at children with various learning difficulties. By creating an emotionally responsive system that adapts to each child's unique needs, the research highlights the importance of personalized learning experiences that go beyond traditional methods.

Looking to the future, the potential for further enhancing this platform is immense. With continued advancements in Augmented Reality (AR) and Virtual Reality (VR), the system could provide even more immersive and interactive learning experiences. AR/VR technologies could be used to create virtual environments where children can practice speaking in realistic contexts, further bridging the gap between virtual and real-world communication. These immersive experiences could also help with spatial awareness and memory retention, two areas often challenging for children with Down syndrome.

Additionally, the integration of more sophisticated natural language processing (NLP) models could allow for even finer-grained analysis of the child's speech patterns, enabling more nuanced feedback. These models could also enable the platform to assess a broader range of phonetic and linguistic challenges, offering even more targeted interventions. Wearable technologies, such as speech-tracking devices or smartwatches, could provide real-time feedback outside of the app, further enhancing the learning experience and offering continuous support.

This study lays a robust foundation for further research and development in the area of speech and language therapy for children with Down syndrome. It opens new avenues for incorporating technology into special education, with the potential to revolutionize the way speech therapy is delivered. By combining intelligent algorithms, emotion-aware systems, and engaging learning strategies, this platform could pave the way for more inclusive, accessible, and effective language development tools. As a result, the approach offers not only a promising therapeutic solution but also an important contribution to the broader educational landscape for children with cognitive and speech impairments. The outcomes of this research provide valuable insights into the role of personalized, technology-driven interventions in improving the communication skills of children with Down syndrome, supporting their academic and social success.

5 REFERENCES

[1] "Facts, Myths, & Truths About Down Syndrome | National Down Syndrome Society (NDSS)." Accessed: Apr. 07, 2025. [Online]. Available: https://ndss.org/myths-truths

[2] "Current perspectives on Down syndrome: Selected medical and social issues - Crissman - 2006 - American Journal of Medical Genetics Part C: Seminars in Medical Genetics - Wiley Online Library." Accessed: Apr. 07, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/ajmg.c.30099

[3] "About Down Syndrome | National Down Syndrome Society (NDSS)." Accessed: Apr. 07, 2025. [Online]. Available: https://ndss.org/about

[4] D. J. Fidler, D. E. Most, and M. M. Guiberson, "Neuropsychological correlates of word identification in Down syndrome," *Research in Developmental Disabilities*, vol. 26, no. 5, pp. 487–501, Sep. 2005, doi: 10.1016/j.ridd.2004.11.007.

[5] S. Buckley, "Language development in children with Down syndrome - Reasons for optimism," *Down Syndrome Research and Practice*, vol. 1, no. 1, pp. 3–9, Jan. 1993, doi: 10.3104/reviews.5.

[6] "Behavioral phenotype of individuals with Down syndrome - Chapman - 2000 - Mental Retardation and Developmental Disabilities Research Reviews - Wiley Online Library." Accessed: Apr. 07, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/1098-2779(2000)6:2%3C84::AID-MRDD2%3E3.0.CO;2-P

[7] "Topics in Language Disorders." Accessed: Apr. 07, 2025. [Online]. Available: https://journals.lww.com/topicsinlanguagedisorders/abstract/2004/04000/communication_skills___the_foundation_for.7.aspx

[8] "Effects of Dose Frequency of Early Communication Intervention in Young Children With and Without Down Syndrome | American Journal on Intellectual and Developmental Disabilities." Accessed: Apr. 07, 2025. [Online]. Available: https://meridian.allenpress.com/ajidd/article-abstract/119/1/17/1185/Effects-of-Dose-Frequency-of-Early-Communication?redirectedFrom=fulltext

[9] "The Efficacy of Treatment for Children With Developmental Speech and Language Delay/Disorder: A Meta-Analysis: Journal of Speech, Language, and Hearing Research: Vol 47, No 4." Accessed: Apr. 07, 2025. [Online]. Available: https://pubs.asha.org/doi/10.1044/1092-4388%282004/069%29

[10] "(PDF) Speech and language therapy for children with Down Syndrome; a systematic literature_review_Feb_2012," ResearchGate. Accessed: Apr. 07, 2025. [Online]. Available: https://www.researchgate.net/publication/236208491_Speech_and_language_therapy_for_children_with_Down_Syndrome_a_systematic_literature_review_Feb_2012

[11] "Modeless Magnetic Bearing System Tracking Using an Adaptive Fuzzy Hermite Neural Network Method | IEEE Journals & Magazine | IEEE Xplore." Accessed: Apr. 07, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/8672598

[12] C. J. Lemons, S. A. King, K. A. Davidson, C. S. Puranik, S. Al Otaiba, and D. J. Fidler, "Personalized reading intervention for children with Down syndrome," *Journal of School Psychology*, vol. 66, pp. 67–84, Feb. 2018, doi: 10.1016/j.jsp.2017.07.006.

[13] "TinyEYE | Online Therapy for Schools | Speech & Occupational," TinyEYE Online Speech Therapy for Kids. Accessed: Apr. 07, 2025. [Online]. Available: https://tinyeye.com/

[14] A. McDuffie and P. Yoder, "Types of Parent Verbal Responsiveness That Predict Language in Young Children With Autism Spectrum Disorder," Journal of Speech, Language, and Hearing Research, vol. 53, no. 4, pp. 1026–1039, Aug. 2010, doi: 10.1044/1092-4388(2009/09-0023).

[15] E. Mason-Apps, V. Stojanovik, C. Houston-Price, and S. Buckley, "Longitudinal predictors of early language in infants with Down syndrome: A preliminary study," *Research in Developmental Disabilities*, vol. 81, pp. 37–51, Oct. 2018, doi: 10.1016/j.ridd.2017.12.021.

[16] "Gestures and Words in Early Development of Children With Down Syndrome | Journal of Speech, Language, and Hearing Research." Accessed: Apr. 07, 2025. [Online]. Available: https://pubs.asha.org/doi/10.1044/jslhr.4105.1125

[17] "Communication intervention for individuals with Down syndrome: Systematic review and meta-analysis: Developmental Neurorehabilitation: Vol 21 , No 1 - Get Access." Accessed: Apr. 07, 2025. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/17518423.2016.1212947

[18] "Can individuals with Down syndrome improve their grammar? - Sepúlveda - 2013 - International Journal of Language & Communication Disorders - Wiley Online Library." Accessed: Apr. 07, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/1460-6984.12002

[19] "Integrated speech and phonological awareness intervention for pre-school children with Down syndrome - van Bysterveldt - 2010 - International Journal of Language & Communication Disorders - Wiley Online Library." Accessed: Apr. 07, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.3109/13682820903003514

[20] K.-A. B. Næss, "Development of phonological awareness in Down syndrome: A meta-analysis and empirical study," *Developmental Psychology*, vol. 52, no. 2, pp. 177–190, 2016, doi: 10.1037/a0039840.

[21] K. Carlstedt, G. Henningsson, and G. Dahllöf, “A four-year longitudinal study of palatal plate therapy in children with Down syndrome: effects on oral motor function, articulation and communication preferences,” *Acta Odontologica Scandinavica*, vol. 61, no. 1, Art. no. 1, Jan. 2003, doi: 10.1080/ode.61.1.39.46.

[23] S. Buckley, “Developing the speech and language skills of teenagers with Down syndrome,” *Down Syndrome Research and Practice*, vol. 1, no. 2, pp. 63–71, Jan. 1993, doi: 10.3104/reports.12.

[24] R. Arias-Marreros, K. Nalvarte-Dionisio, and L. Andrade-Arenas, “Design of a Mobile Application for the Learning of People with Down Syndrome through Interactive Games,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 11, Art. no. 11, 57/01 2020, doi: 10.14569/IJACSA.2020.0111187.

[25] “Mobile Speech Therapy Application Using Speech Processing for Intellectually Disabled | Journal of Modern Rehabilitation.” Accessed: Apr. 07, 2025. [Online]. Available: https://jmr.tums.ac.ir/index.php/jmr/article/view/133

[26] “(PDF) Teaching Children with Down Syndrome Pronunciations Using Speech Recognition,” ResearchGate. Accessed: Apr. 07, 2025. [Online]. Available: https://www.researchgate.net/publication/206005894_Teaching_Children_with_Down_Syndrome_Pronunciations_Using_Speech_Recognition

[27] Md. Sahidullah and G. Saha, “Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition,” *Speech Communication*, vol. 54, no. 4, pp. 543–565, May 2012, doi: 10.1016/j.specom.2011.11.004.

6 APPENDICES

Survey Link - https://forms.gle/QAtP4zw7gkEtuYQ47

Plagiarism Report -

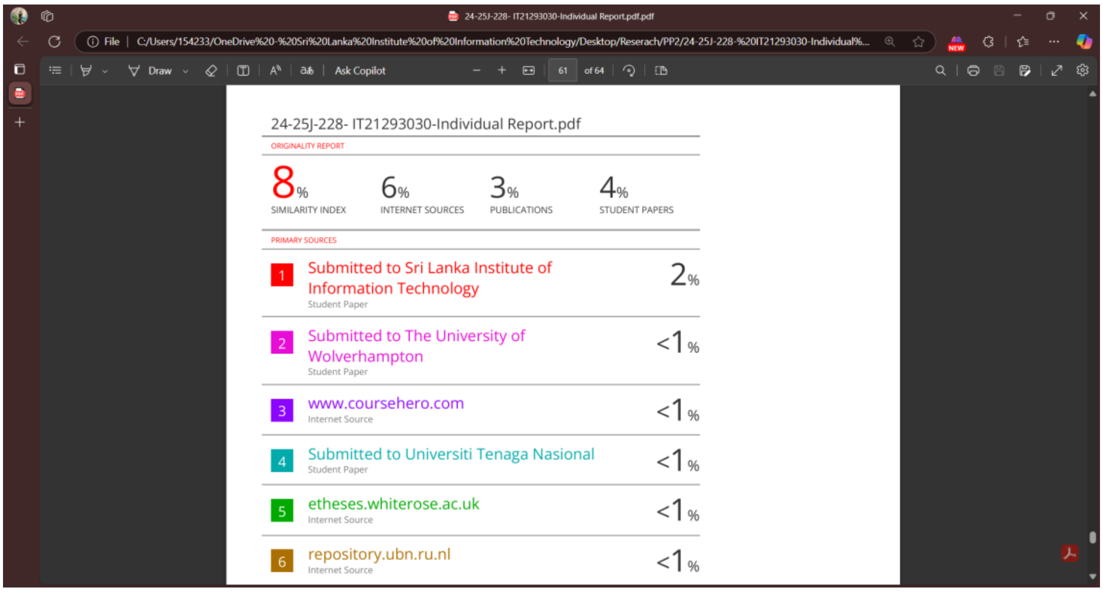

Figure 33 - Plagiarism Report