Interactive, visual learning-based tool for hearing impaired children to improve language skills

Udbhasa M M S
Faculty of Computing
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
it19188929@my.sliit.lk

Lelkada L L P S M
Faculty of Computing
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
it19001708@my.sliit.lk

Priyanka P D M K
Faculty of Computing
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
it19954974@my.sliit.lk

Senarathna B W E K
Faculty of Computing
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
it19142838@my.sliit.lk

Janaka L. Wijekoon
Faculty of Computing
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
janaka.w@sliit.lk

Samitha Vidhanaarachchi
Faculty of Computing
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
samitha.v@sliit.lk

***Abstract*—Hearing impairment is a common condition that affects millions of people worldwide, with approximately 1 in 1000 newborns experiencing some degree of hearing loss. This condition can significantly impact a person's quality of life, causing communication difficulties and social isolation. Early childhood hearing impairment can pose significant challenges to a child's cognitive development, making it difficult for them to learn in a traditional learning environment. With advanced treatment options available, hearing-impaired children can now hear, but they still face challenges in language development due to a lack of auditory and cognitive stimulation. Delayed language development can lead to lower academic achievement and social isolation. Therefore, there is a need for effective learning tools to aid hearing-impaired children in learning their first language. This paper proposes visual-based and interactive learning tools as an effective method for enhancing the learning experience and engagement of hearing-impaired children. The proposed method comprises of a mobile application which is proven to help with 9 out of 10 children(90%) in their early childhood.**

Keywords—Hearing impaired children, Learning tool, Language development, Cognitive development.

I. INTRODUCTION

Hearing impairment is one of the most common conditions that affects millions of people worldwide. According to the World Health Organization (WHO), approximately 1 in 1000 newborns have hearing loss at degree of mild to profound. This condition has a significant impact on a person's quality of life as it causes communication difficulties and social isolation.

The learning journey of a child begins from the moment they are born. Cognitive development during early childhood lays the foundation for their future academic success and lifelong learning. Effective communication is the key for language development and social - emotional development of a child. Accordingly, being hearing impaired cause significant effect on child's cognitive development. Hearing impairment is a condition of hearing loss from mild to profound. As this condition affects a child's ability to hear, it makes it difficult for a child to learn in a traditional learning environment.

With the significant advancement in medical technology of the past decade, there are many advanced treatment options are available for hearing impaired people. Even though hearing impaired children are fortune to get the ability to hear with those advanced treatments, as "Poverty of the stimulus argument" by Noam Chomsky explains the condition where lack of auditory simulation makes mastering a language difficult for a hearing-impaired child.

Learning the first language in early childhood lays the foundation for cognitive development. According to research conducted on the cognitive development of a child, it is crucial for a child to learn his first language by the age of 5. Linguistic skills significantly impact a child's memory development, attention, and problem-solving skills that are directly connected with academic success. Delayed language development of a hearing-impaired child causes lower level of academic achievement and social isolation. Therefore, there is a need of effective learning tools for hearing impaired child to learn his first language.

Visual-based and interactive learning tool is an effective method for enhancing the learning experience and engagement of hearing-impaired children. Compared to the traditional learning tools, visual learning tools can be used for providing a more dynamic and interactive learning experience.

II. LITERATURE REVIEW

Visual-based and interactive learning tool is an effective method for enhancing the learning experience and engagement of hearing-impaired children. Compared to the traditional learning tools, visual learning tools can be used for providing a more dynamic and interactive learning experience.

Language development is crucial for a child's cognitive growth, and it plays a vital role in communication, problem-solving, and relationship-building abilities [1]. However, hearing-impaired children face challenges in learning sounds and gestures, which can impact their ability to express themselves, understand others, and develop cognitive abilities like memory, reasoning, and visualization [2][3].

Language development has five stages [4], from pre-production to advanced fluency, but hearing-impaired children may need to learn new words and improve their vocabulary to develop cognitive abilities fully. To address

this issue, gamifying language learning benefited hearing-impaired children by allowing them to learn at their own pace, providing engaging content, and a customized learning experience [5][6]. Gamification encouraged specific behaviors, engaged children effectively, and helped them track their progress [7]. Using reinforcement learning and Q-learning algorithms analyzed factors such as response time, accuracy rate, and quiz difficulty to determine the learner's initial status and progress.

Online learning tools that gamify language learning improved teaching methods, suggest suitable materials, and enhance a learner's growth [8][9]. With hearing-impaired children facing challenges in accessing audiology centers for training, online learning tools can provide a convenient and cost-effective solution. Additionally, with the shift towards online learning accelerated by the COVID-19 pandemic, academic leaders consider online education to be just as good, if not better than traditional classroom-based educational offerings [4]. Overall, an online learning tool that gamifies language learning helped hearing-impaired children develop language skills and cognitive abilities, leading to improved school performance, IQ, and life trajectory.

The cognitive development of a hearing-impaired child is adversely affected in specific areas like language and audible memory, despite being capable of hearing sounds with high dynamic range and distinguishing sounds of words and sentences. Language acquisition is negatively impacted by hearing impairment, and the effects of language development delay persist throughout their lives [10]. Thus, hearing-impaired children require special care and attention to improve their linguistic skills, and alternative methods and techniques are necessary to raise them to have a normal life [11]. Brain plasticity in early childhood supports rapid language acquisition, and it is crucial to learn languages during the first 4.5 years [12]. Unfortunately, the misconception that hearing aids solve everything often worsens the situation, as parents do not put in further effort and wait for the child to speak miraculously at school [11]. Since hearing-impaired children do not get exposed to spoken language, their vocabulary is limited, which heavily affects their education, as education is heavily biased towards verbal languages [13]. A study conducted by the Community Child Health, Royal Children's Hospital, Australia, found a relationship between the age of detection of hearing impairment and education outcomes [14]. The study involved 132 students aged 7-8 years who underwent a three-hour assessment to measure educational outcomes under the supervision of a speech pathologist and a psychologist.

Language acquisition is a crucial aspect of early childhood development, and research suggests that brain plasticity plays a vital role in supporting rapid language acquisition. Children who are exposed to their surroundings on a regular and frequent basis naturally become fluent in their native language by the age of five [15]. However, hearing-impaired children do not get exposed to spoken language, limiting their vocabulary to a few hundred words by the age of five, which heavily affects their education [16].

To study the impact of hearing impairment on education outcomes, researchers from the Community Child Health, Royal Children's Hospital, Australia, conducted a study on 132 students aged 7-8 years [17]. The study aimed to identify the relationship between the age of detection of hearing impairment and education outcomes. The study found that early diagnosis of hearing impairment is crucial for better linguistic abilities in children, regardless of the severity of the impairment.

The study measured educational outcomes using various tests, such as the Clinical Evaluation of Language Fundamentals–Third Edition (CELF), Peabody Picture Vocabulary Test (PPVT – 3rd edition), and the Wechsler Intelligence Scale for Children Third Edition (WISC). The study found that early detection of hearing impairment alone is insufficient to substantiate the long-term benefits of early detection, and a mechanism to acquire language is necessary.

Overall, this research highlights the critical role of early detection and intervention in hearing impairment to support better language acquisition and education outcomes in children.

The education of hearing-impaired students can be negatively impacted by improper parenting and lack of awareness. Early intervention with special education can help these students develop language skills and improve their cognitive abilities. To address this issue, a proposed student-centric learning system has been developed that provides an interactive environment where children can learn from their surroundings and practice words and signs. The system monitors the level of the student and selects the best course material for them. Additionally, measures are included to improve verbal communication using lip exercises and speech therapy.

Previous research has focused on the development of sign language skills with the aid of computer games. A study in Portugal used a web-based game application that detected the user's skeleton and hand gestures to teach sign language to students aged 7-8 [18]. In India, a team of researchers developed a mobile application that teaches the Indian Sign Language (ISL) to hearing-impaired students [19]. The application focused on developing cognitive skills by teaching opposite and contrast words in ISL and by providing simple and comprehensive arithmetic equations.

Object detection is an important component of the proposed system, and the YOLO (You Only Look Once) algorithm was selected for its real-time speed and high performance [20]. The YOLO algorithm has evolved from version 1 to version 5 [21], with improvements including anchor with K-means, multi-scale detection, and flexible control of model size. The YOLOv5 algorithm was found to show higher average precision than YOLOv3 and YOLOv4 [22]. Various subversions of YOLOv5 can be optimized to run on mobile environments, making it suitable for use in the proposed system.

III. METHODOLOGY

This study further divides into four subcomponents as Object exploration, Similar word generation, Lip movement analyzing and Level determination. High level relationships between the components are illustrated in the below figure.

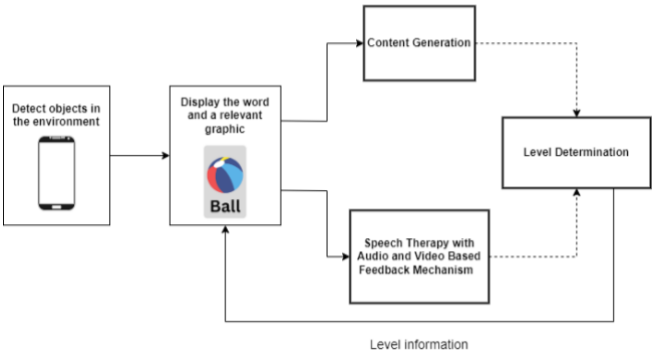


Figure 1. High level system architecture

A. *Object Exploration*

Exposing the child to the learning environment is important in learning through the environment. This is mainly based on the hearing-impaired child detecting an object and identifying the object. Initially the environment around the child is captured by means of the mobile camera and directed into an object detection model. The captured images will be directed to a game where the child is involved in a learning curve to expose the objects in the surroundings. Based on the level of the child the complexity of the objects given to him will differ. Initially 3 letter words will be exposed to the child in the surroundings and gradually increased for more complexities with inclusion of 4 to 5 letter words. Child users are requested to drag and drop the letters in the order of the word they are given. The statistics related to child responses are also monitored at this point.

The objects in the surrounding child environment are tracked using an object detection model. Initially the model was trained using YoloV3, YoloV4 and Yolo MobileNet. The MobileNet Yolo had good GPU performance in mobile but had low precision compared to YOLO V5 and the number of detections per frame showed a significant loss compared to Yolo V5. The usage of high number of pixels (1280px) and new equations YoloV5 shows good precisions when identifying boundaries of detected bounding boxes Yolo V5 made it. Since the child is involved in a real-time learning system, Yolo v5 detection speed (766ms) detected objects in frames at high speed, Yolo v5 algorithm is used to track objects. The model will be trained using general objects in the domestic environments and the objects i.e. cat, dog, apple, pen etc. The detected objects are labeled with very simple class names which are comfortable for child users to pronounce. The detections should also not be too complicated for the child. Direct output from object detections will result in overlapping objects displayed in a single frame a child encounters. This results in issues in the learning experience of the child and to overcome this modular algorithm is used for this purpose. Based on the vocabulary of the child object detection will be limited. Yolo creates bounding boxes around each object and the detections are subjected to filtration based on the learning curve of the child and the complexity of the environment. The IOU (Interest of union) is used to determine the most prominent object detected in the frame. If the overlapping area of 2 objects surpasses the threshold IOU value of 0.6 the object with low exposure is eliminated from the frame. This results in the child focusing on the prominent object in the frame.

The detected objects are subjected to crop segmentation. The segmentation mainly focusses on capturing the limits of the objects which the child captures i.e., color of the object, size of the object. Mask RCNN will be used to crop segment the objects. The ResNet50 + FPN layer will be used to extract the feature maps of corresponding objects. Based on the key features of the object a fixed-size feature map is generated to each object. It is based on the ROI (Regions of interest) The feature maps are subjected to an FCN to pixel wise segment the objects. The masked objects are cropped and displayed to the child to understand the feature boundaries of the object. The crop segmenting model is trained using images samples of the domestic environments in households.

The variety of learning environments causes limitations in the capturing of all objects related to the learning curve of the child. As a remedy an object classification model is introduced. The approach of classification with a minimum of 10 images involves a one-shot classification of objects. In the classification a set of base images is used to generate a similarity score for the target object. The highest similarity score will guarantee the class of the object. The similarity score is generated based on a Siamese neural network (with 4 layers). Similarity score takes a value between 0 and 1 and relative show how much the 2 images are related. When testing the model several sizes for images were tested and optimal results were obtained at 105X105px and based on these values the final Siamese network was trained with a value of 38951745 trainable parameters. The similarity score is generated based on the greyscale pixels of the base image and the target image. With all these values the model takes around 20 minutes to train with the addition of a new object to the model. The training is very fast compared to traditional classification models and the training image count is also considerably low. It is recommended to use plain backgrounds (solid color background) for the classifications to achieve higher accuracies.

A hybrid model of YOLOv5 and Siamese One-Shot learning model will be used to run simultaneously for both local object detections and for specific newly introduced objects. The objection model will be an in-device model and the one-shot classification model will be hosted on a remote server.

B. *Contextually Similar Word Generation*

This study aims to generate contextually similar words based on the acquired vocabulary. Word generation sub component takes explored words from the Object exploration sub component as an input and generate contextually similar words based on the input word. Following figure illustrate the flow of contextually similar word generation.

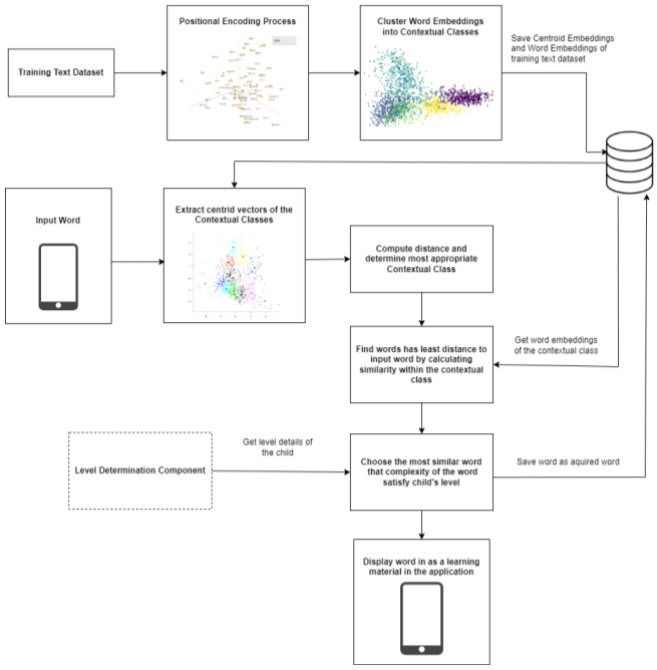


Figure 2. Flow of similar word generation

As the intended audience of this application is children, it is important to use customized dataset for model training to ensure the child friendliness of the language model. The collected training corpus contains nearly 80 million of tokens extracted from public datasets, children story books. Using the training corpus, a language model is created in order to support different sub processes in word generation subcomponent. Each word of the corpus is encoded a fixed length vector using a positional encoding process that provide information about order of the tokens in a sequence of tokens. Positional encoding enables the ability to distinguish different permutations a token can have. Resulted word embeddings are dense and low-dimensional vector representations which capture the semantic relationships between words. Vector size of the word embeddings used in this study contains 768-dimensional embeddings. Number of dimensions are decided considering the performance of the downstream tasks and computational costs.

Resulted embeddings are clustered into n number of groups, using k-means clustering technique. Goal of word embedding clustering is to identify groups of similar words together based on their contextual meanings. K-means clustering algorithm is adapted for word embeddings by calculating distance between word embeddings as a similarity measurement. This approach assign a word to a cluster where the sum of squared distances between each word embedding and its assigned cluster centroid is minimal.

$$\textbf{minimize: } J = \sum_{i=1}^{n} \|w_i - c_{j(i)}\|^2$$
$$\textbf{subject to: } j(i) = \text{argmin}_j\, d(w_i, c_j)$$

where,
n = number of word embeddings
w_i = word embedding
c_j(i) = centroid assigned to a given w_i
j(i) = index of the centroid closest to w_i

As k-means clustering is sensitive to the initial centroids, algorithm is run multiple times with different initial centroids in order to select the set of clusters with lowest error. Preliminary steps are completed with choosing most appropriate centroids.

When a object is detected and displayed by the Object Exploration sub component, the word is recorded as a acquired word in the database. Acquired vocabulary serves as the input for word generation subcomponent. initially extracts the word embedding of the input word and subsequently determines the context class by measuring its distance against a pre-defined set of centroids. Context class of that contains the centroid that has least distance to the input word is considered as the context class of the input word. To compute the distance, Euclidean distance is used that can be represented as,

$$\textbf{d = sqrt(sum(xi - yi)^2)}$$

Within the context class of the input word, most similar contextual word is extracted by computing the Euclidean distance against the words in the same contextual class. Other than the similarity for a given word, recommended word complexity from the Level Determination subcomponent also considered. Chosen contextually similar word is displayed in a gamified environment as the final step of the contextual Similar Word Generation subcomponent.

C. Speech therapy using voice and lip-reading based feedback mechanism

The speech therapy component involves a feedback mechanism which captures a child's pronunciation using audio and video. A composite error rate was calculated to detect pronunciation mistakes for the displayed word.

1) Speech detection using audio

A deep learning algorithm called "Deep Search" is used for on device word classification using audio. The implementation of the model involves several key steps. First, the raw audio data is preprocessed to extract spectrogram features, which are then fed into the model. The model then uses multiple layers of LSTM units to learn representations of the input features, with each layer adding additional complexity and abstraction.

To further improve the accuracy of the model, the authors propose a novel technique called Connectionist Temporal Classification (CTC) loss. The loss function of this algorithm allows to handle variable-length inputs and outputs, which is critical for accurately transcribing speech. Additionally, the authors employ a technique called beam search decoding, which helps to further refine the model's output by exploring multiple possible transcriptions and selecting the most likely one.

The RNN model has a simpler structure than similar models in the literature [14]. This is because we have limited ourselves to a single recurrent layer, which is difficult to parallelize, and have excluded the use of Long-Short-Term-Memory (LSTM) circuits. LSTM cells are disadvantaged by their need to compute and store multiple gating neuron responses at each step, which can become a computational bottleneck during forward and backward recurrences. By using a homogeneous model, computation of recurrent

activations has optimized to be as efficient as possible. This involves performing only a few highly optimized BLAS operations on the GPU and applying a single point-wise nonlinearity to compute the ReLU outputs.

2) Video based lip-reading

Lip reading is the main mode of validating speech when noise is present. The input from the camera captures the facial landmarks and the lip movements.

BlazeFace model employs an enhanced network architecture derived from MobileNet. The original authors observe that the computation time for a 3x3 depthwise convolution of a 56x56x128 tensor is only 0.07ms on iPhoneX cpu, while the subsequent 1x1 convolution from 128 to 128 channels takes 4.3× longer at 0.3ms. Based on this, they suggest that enlarging the kernel size of the depthwise convolution is a relatively inexpensive operation. Therefore, the authors propose replacing the 3x3 depthwise convolution with a 5x5 depthwise convolution, which reduces the model's depth and improves processing speed.

$$[u_t, r_t]^T = \text{sigm}(W_z z_t + W_h h_{t-1} + b_g)$$

Where, z := {z1 , . . , zT } is the input sequence to the RNN.

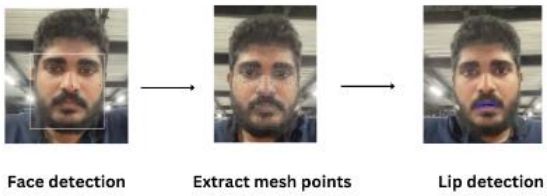


Figure 3. Image processing pipeline

There are several of lipreading datasets available such as AVICar, AVLetters, AVLetters2, as documented in previous research (Zhou et al., 2014; Chung & Zisserman, 2016a). However, most of these datasets consist of only single words or are too small in size. GRID corpus dataset (Cooke et al., 2006) , includes both audio and video recordings of 34 speakers producing 1000 sentences from each person, resulting in a total of 28 hours of data comprising 34,000 sentences.

The proposed model is based on a 3D convolutional neural network that takes in video frames as input and outputs text at the sentence level. The model is trained on a large-scale dataset of audio-visual sentences and achieves state-of-the-art results in lipreading accuracy. The authors propose a novel approach to integrating audio and visual information by fusing the two streams in the later layers of the network. LIPNET model outperforms existing lipreading models and holds promise for various applications, such as improving speech recognition systems for noisy environments.

The audio component's score and the Lipnet's score is individually calculated and later they were aggregated using a weighted average. The results were then forwarded to level determination and other modules for their use.

$$Weighted\ Average = \frac{\sum wx}{\sum w}$$

where,

w = the weight for each score.
x = the value of each score.

D. Initial Status Determination and Progress Monitoring

The component consists of two main subcomponents, which aim to determine the initial status of the user and monitor their progress. The first task is to use a quiz to evaluate the user's initial status dynamically. The quiz will display simple words from 5 different levels. Then it analyzes the response accuracy, response time, and final quiz difficulty using a random forest algorithm to determine the user's initial status, which will be saved in the database. The second subcomponent involves monitoring the user's progress using the levels they obtain every time the app is visited. It rewards the child using a reward mechanism to motivate the user to perform well.

As this solution has been built with the collaboration of CEHIC the level determination learning materials could be adopted. The method that is primarily used for determining the level is ‘’by the number of letters in a word”. Furthermore, this evaluates the knowledge on basic words as three letter words, four letter words, five letter words and complex words which has more than five letters or a combination of two words.

Three Letter Words	Four Letter Words	Five Letter Words	Compound Words
Cat	Bike	Spoon	Airplane
Bat	Kite	Board	Bathtub
Rat	Tree	Cobra	Cupcake
Mat	Coin	Knife	Dustbin
Dog	Pond	Skirt	Lipstick
Car	Gift	Wheel	Milkshake
Hat	Vase	Shirt	Staircase

Each question is displayed along with a pronouncing voice clip of the word playing in the background.

Regression analysis is a statistical method that enables to understand how the value of a dependent variable changes in relation to an independent variable.[1] The component uses a model implemented with the regression algorithm to determine the level of the user based on the response accuracy, response time, and final quiz difficulty. Random forest regression is a regression algorithm that models the relationship between a dependent and independent variable as an nth degree polynomial.[2] This method is used when data does not have linear relationships between each other and there are multiple independent variables. [3] The personalized data of the user does not have linear relationships with each other, so this method can be used to evaluate the progress level of the user.

The gamified e-learning system will implement these methods to motivate the user to perform well. Overall, this approach is an alternative method for personalized systems with challenging objectives.

The gamified solution has a quiz which has fifteen questions in total in five different levels. Three questions will be randomly selected from each level for the quiz and those will be displayed. The child must select the accurate image then drag and drop that into the basket. Fifteen seconds are given to answer one question. The audio clip pronouncing the word can be played any number of times within those fifteen seconds. If the child answers the question incorrectly they can still try again until the fifteen seconds are over. The next question will be displayed as soon as the timer reaches zero. There is an option available to skip a question and navigate into the next question as well.

The result of the quiz is taken into the random forest algorithm as a processed data set. The data taken are answer accuracy, response time and difficulty of the question which will then be used to determine the initial level of the user.

The dataset used for the training of random forest algorithm consisted of accuracy of answers given by the child for each question in the quiz. In order to acquire the difficulty of the question different points were given for each level of question. Those points are allocated if the answer was right and zero is allocated if the answer is incorrect.

Column	Points Given
Difficulty Level 1 Question 1	1
Difficulty Level 1 Question 2	1
Difficulty Level 1 Question 3	1
Difficulty Level 2 Question 1	2
Difficulty Level 2 Question 2	2
Difficulty Level 2 Question 3	2
Difficulty Level 3 Question 1	3
Difficulty Level 3 Question 2	3
Difficulty Level 3 Question 3	3
Difficulty Level 4 Question 1	4
Difficulty Level 4 Question 2	4
Difficulty Level 4 Question 3	4
Difficulty Level 5 Question 1	5
Difficulty Level 5 Question 2	5
Difficulty Level 5 Question 3	5
Total Points	Sum of Points Earned
Time Spent	Total Time Spent for Quiz

The random forest algorithm first processes the dataset and comes up with a correlation matrix. The correlation matrix is built with analyzing the variables in rows and columns. Each cell in the matrix contains the coefficient of the correlation of variables. [4] The variables which are impacting heavily on the level of the child is taken according to this and then data is preprocessed and split as train and test data. In the developed solution 20% of data are taken for testing. Then the model is training along with the hyperparameters which are n_estimator, max depth and min_samples_leaf. N_estimator specifies the number of trees[4], max_depth specifies the number of splits that each tree is allowed to make and min_samples_leaf specifies the minimal number of samples required at the leaf. After the training of data, the results that give maximum accuracy for detecting the level of the user are taken and passed to get the maximum accurate level of the child.

IV. Results

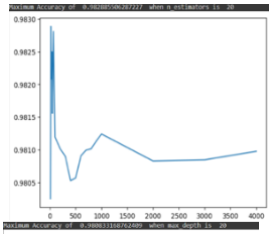
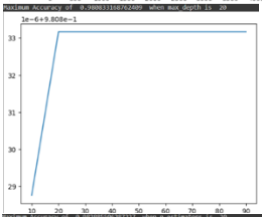
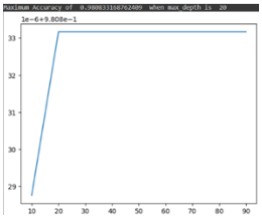
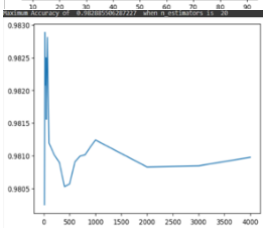
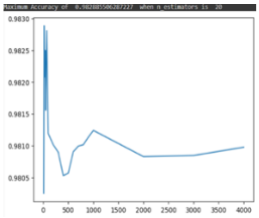
For the speech therapy component Grid corpus is used as the primary training data source. In this task, the goal is to transcribe the spoken sentences in the audio recordings while also identifying the corresponding visual features of the speaker's mouth and face movements. Common metrics used to evaluate the performance of audio-visual speech recognition models on the GRID corpus include word error rate (WER), which measures the proportion of transcribed words that differ from the ground truth, and frame error rate (FER), which measures the proportion of incorrectly identified visual frames. Other metrics such as precision, recall, and F1 score may also be used depending on the specific research question and evaluation criteria. Overall, the GRID corpus provides a valuable resource for researchers working on audio-visual speech recognition and related tasks.

$$WER = (S + D + I) / N$$

where , S is the number of substitution errors, D is the number of deletion errors, I is the number of insertion errors, and N is the total number of words in the reference transcription.

	CER	WER
Hearing-Impaired Person (avg)		47.7%
LSTM model	38.4%	52.8%
Ours	6.5%	11.5%

Accuracy, a fundamental metric in evaluating regression models, continues to be the cornerstone of performance measurement. It provides an intuitive and straightforward way to assess the effectiveness of models, and as such, remains a widely used estimate. In the developed solution, random forest algorithm model is tailored to specific use case of determining the initial level of the user. The split train-test dataset is into an 8:2 ratio. The accuracy of model is determined by the n_estimator, max depth and min sample leaf. These results offer valuable insights into the performance of the model, and can inform future research and real-world applications in this domain.

As per the results if n estimator is taken as 20, max depth is taken as 20 and min sample leaf is taken as 1, the random forest algorithm can give 0.980833 of accuracy which is 98%.

Following matrices are used in order to assess the performance of word embeddings clustering.

Silhouette coefficient

$$s=(b-a)/\max(b-a)$$

Where ,

a = average distance to sample against all the other points within the cluster

b = average distance to sample against all the other points closest neighbor the cluster

V. Discussion

It was observed that children show a considerable interest in learning new words through the learning environment created through mobile phone. With mechanism to track their lip movements and pronunciation increases their learning experience. The language used by the child is also renewed by providing new content/words based on the content/words he has already learnt. The monitoring of the hearing-impaired child learning progress shows the child's level of learning. Based on the learning the level up and level down process can be handled.

VI. Conclusion and Future Work

Using a mobile app with gamified tasks and explorative learning this paper proposes a method which can improve language and cognitive abilities of a hearing-impaired child. It was observed that all the models perform with performance with better than 84% accuracy. The application should be used in the field for many years to gather its real impact on the students.

The above research work is done based on English language and further extension can be taken to do the enhanced the visual learning to other languages. Currently the models are compatible with mobile versions and further enhancement related to increase the compatibility of mobile phones can be done. The learning process is currently done with the supervision of a teacher on premises. But further improvements can be made to accommodate remote learning.

Acknowledgment

We would like to express our sincere gratitude to our supervisor, Dr. Janaka Wijekoon, for his unwavering guidance, support, and valuable insights throughout the research process. His commitment, expertise, and patience have been instrumental in shaping this work. We would also like to extend our appreciation to our co-supervisor, Mr. Samitha Vidhanaarachchi, for his invaluable feedback, technical support, and encouragement. His dedication and expertise have been crucial in ensuring the success of this project.

Furthermore, we would like to thank Rev. Sr. Greta Nalawatta from the Centre For Education Of Hearing Impaired Children (CEHIC) for her insightful comments and recommendations. Her expertise in the field has been instrumental in shaping the direction of our research.

We would also like to acknowledge the contributions of our colleagues who have provided us with support and encouragement throughout the research process.

References

[1] Christine Yoshinaga-Itano, Allison L. Sedey , Craig A. Mason, Mallene Wiggin, Winnie Chung, "Early Hearing Detection and Vocabulary of Children with Hearing Loss", August, 2017

[2] Rufsvold, Ronda. "The impact of language input on deaf and hard-of-hearing preschool children who use listening and spoken language" Columbia University, 2018.

[3] Fidaa Almomani1, Murad O. Al-momani , Soha Garadat, Safa Alqudah, Manal Kassab, Shereen Hamadneh , Grant Rauterkus and Richard Gans, "Cognitive functioning in Deaf children using Cochlear implants", 2021

[4] Computer Systems Institute, "The Five Stages of Learning a Language", Available: https://www.csinow.edu/blog/five-stages-learning-new-language/

[5] Boyan Bontchev and Dessislava Vassileva, "Educational Quiz Board Games for Adaptive E-Learning", Sofia University "St. Kliment Ohridski", June 2010

[6] Jan L. Plass, Bruce D.Homer, Charles K. Kinzer, "Foundations of Game-Based Learning", New York University,2015

[7] Shabnam Mohamed Aslam, Abdul Khader Jilani, Jabeen Sultana, Laila Almutairi, "Feature Evaluation of Emerging E-Learning Systems Using Machine Learning: An Extensive Survey", May 5, 2021.

[8] Victoria from Teach Smarter, "Benefits of Monitoring Student's progress in the Classroom". Available: https://www.teachstarter.com/us/blog/4-benefits-monitoring-student-progress-classroomus/

[9] Erica Lembke, “Supporting Teachers Who Are Implementing Student Progress Monitoring: A Guide for Administrators 2006 Summer Institute on Student Progress Monitoring” 2006

[10] Meadow, Kathryn P. “Early Manual Communication in Relation to the Deaf Child’s Intellectual, Social, and Communicative Functioning.” Journal of Deaf Studies and Deaf Education, vol. 10, no. 4, 2005, pp. 321–29. JSTOR, http://www.jstor.org/stable/42658773. Accessed 9 Oct. 2022.

[11] McNeill, D. “The capacity for language acquisition. In Research on behavioral aspects of deafness”, National Research Conference on Behavioral Aspects of Deafness, New Orleans, LA. Vocational Rehabilitation Administration, 1965

[12] T. Humphries et al., "Language acquisition for deaf children: Reducing the harms of zero tolerance to the use of alternative approaches", Harm Reduction Journal, vol. 9, no. 1, 2012. Available: 10.1186/1477-7517-9-16 [Accessed 8 October 2022].

[13] S. Dixon, "Two Years: Language Leaps**The chapter is dedicated to Elizabeth Bates, Ph.D. (1947–2003), for her early contribution to this work and for her studies of child language worldwide.", Encounters with Children, pp. 382-409, 2006. Available: 10.1016/b0-32-302915-9/50020-3 [Accessed 8 October 2022].

[14] L. Zauche, T. Thul, A. Mahoney and J. Stapel-Wax, "Influence of language nutrition on children’s language and cognitive development: An integrated review", Early Childhood Research Quarterly, vol. 36, pp. 318-333, 2016. Available: 10.1016/j.ecresq.2016.01.015.

[15] T. Humphries et al., "Language acquisition for deaf children: Reducing the harms of zero tolerance to the use of alternative approaches", Harm Reduction Journal, vol. 9, no. 1, 2012. Available: 10.1186/1477-7517-9-16 [Accessed 8 October 2022].

[16] S. Dixon, "Two Years: Language Leaps**The chapter is dedicated to Elizabeth Bates, Ph.D. (1947–2003), for her early contribution to this work and for her studies of child language worldwide.", Encounters with Children, pp. 382-409, 2006. Available: 10.1016/b0-32-302915-9/50020-3 [Accessed 8 October 2022].

[17] M. Wake, "Hearing impairment: a population study of age at diagnosis, severity, and language outcomes at 7-8 years", Archives of Disease in Childhood, vol. 90, no. 3, pp. 238-244, 2005. Available: 10.1136/adc.2003.039354

[18] Filomena Soares, João Sena Esteves, Vitor Carvalho, Gil Lopes, Fábio Barbosa, Patrícia Ribeiro, “Development of a serious game for Portuguese Sign Language” in 2015 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)

[19] Sandipa Roy, Arpan K Maiti, Indira Ghosh, Indranil Chatterjee, Gopal K Basak, Kuntal Ghosh, “An app based unified approach to enhance language comprehension and mathematical reasoning ability of the hearing-impaired using contrast words”

[20] Joseph Redmon∗, Santosh Divvala∗†, Ross Girshick, Ali Farhadi∗†, “You Only Look Once: Unified, Real-Time Object Detection”, in 2016 IEEE Conference on Computer Vision and Pattern Recognition

[21] Peiyuan Jiang, Daji Ergu*, Fangyao Liu, Ying Cai, Bo Ma, “A Review of Yolo Algorithm Developments”, in The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 & 2021)

[22] Upesh Nepal and Hossein Eslamiat, “Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs”