An Interactive Machine Learning Approach for Improving Babble English Pronunciation

Ansari Almas Mubeen[1], Kalekar Sudarshan Ajinath[2], Kapse Om Nilesh[3], Tajane Harshad Kailas[4], Dr. Rokade Prakash Pandharinath[5]

SND College of Engineering & Research Center, Savitribai Phule Pune University

Abstract

This paper introduces a Pronunciation Checker, which uses deep learning techniques to improve users' speaking skills.The Babble English project focuses on advancing English proficiency among non-native speakers by enhancing pronunciation, fluency, and natural language use. Through an interactive platform, Babble English provides a structured environment for learners to practice spoken English in real-time, utilizing AI-driven pronunciation feedback and personalized learning paths. This project emphasizes the improvement of articulation, intonation, and phonetic accuracy, crucial for clear communication. Designed for beginner to intermediate learners, Babble English leverages adaptive learning technologies to analyze speaking patterns and offer targeted exercises, helping users develop confidence and natural fluency in English conversations. The system requires users to register and log in to access a selection of randomly generated paragraphs. Upon reading the paragraphs aloud, the system employs advanced speech recognition algorithms to analyze the user's pronunciation, identifying any mispronounced words. A detailed report is generated, highlighting specific words that were pronounced incorrectly and providing their meanings along with phonetic pronunciations for correction. Additionally, the system tracks the user's progress, presenting a percentage score reflecting their pronunciation accuracy for each paragraph. Users also have the option to directly record sentences, which the system evaluates for correctness. This interactive tool aims to support language learners by providing immediate feedback and fostering continuous improvement in their pronunciation.

Keywords: Pronunciation Checker, Deep Learning, Speech Recognition, Mispronounced Words, Phonetic Pronunciation, Pronunciation Accuracy, Language Learning, Continuous Improvement, User Progress Tracking, Interactive Tool

INTRODUCTION

The Pronunciation Checker is an innovative tool designed to assist users in enhancing their speaking skills by leveraging deep learning techniques. For language learners, achieving pronunciation accuracy is often a critical and challenging aspect of fluency. As noted by SeeminNiclors and Rafal John (2021), the journey to language mastery frequently encounters obstacles in the form of mispronunciations, which can affect communication and confidence. This system provides a practical solution by offering real-time feedback on spoken language, identifying errors in pronunciation, and suggesting corrective phonetic modifications to help learners overcome these challenges.

Upon registering and logging into the system, users are presented with randomly generated paragraphs to read aloud, which introduces an element of unpredictability to test various vocabulary and sentence structures. Utilizing advanced speech recognition algorithms, the system evaluates the user's spoken input

with precision, analyzing the audio for discrepancies in pronunciation. As highlighted by Gorge Jems, Clea Decker, and Cloe (2021), such AI-driven assessments are invaluable in pinpointing specific words that were mispronounced. A comprehensive report is then generated, detailing not only the mispronounced words but also their meanings and correct phonetic pronunciations. This level of analysis allows users to understand their errors contextually, ensuring that they grasp both the meaning and the proper enunciation of the words.

Building on this feedback loop, the system tracks users' progress over time, compiling performance data into a percentage score that reflects their pronunciation accuracy for each paragraph. This score helps learners visualize their improvement and areas requiring further practice. Furthermore, users have the flexibility to record individual sentences for more focused, targeted feedback, addressing specific areas where they struggle.

Beyond the functionality of error correction, the Pronunciation Checker offers an immersive and interactive experience for language learners. As noted by Allena and TrexiShaen (2020), personalized, immediate feedback accelerates learning and builds the confidence needed for fluent speaking. By encouraging consistent practice and providing detailed, AI-powered insights, this tool fosters continuous improvement and helps learners overcome hurdles that may have otherwise hindered their progress. A computing engine tailored for the modern learning environment, the system not only facilitates accurate pronunciation but also supports learners in developing confidence and fluency in their spoken language (2020).

LITERATURE SURVEY

1. AlexanderAntonov, ”Pitfalls of Machine Learning Methods in Smart Grids: A Legal Perspective,”[1] 2021 - This paper is a visual tool in form of a table, meant to guide prosumers, utility, technology and energy service providers. It shows the areas that need increased attention when dealing with specific prosumer concerns as identified in the technical literature.

2. Ronghuai Huang,” A Computing Engine for the New Generation of Learn- ing Environments,”[2] 2020 aˆ To help computer scientists and ICT developers to create technical solutions that can better support new learning environments, we proposed a computing framework, aˆSmart Learning Engine (SLEnG).ˆa Its definition, functions, architecture, and computing process were described. The application ex- amples with the initial results were reported.

3. Mohammad Ali Akour, ”Developing a Virtual Smart Total Learning Environment for Future Teaching-Learning System,”[3] 2020 - The proposed framework is focused on transforming the learning experience into two possible ways like online and on-campus learning through groundbreaking agile methodologies. The new interfaces for learners will be included like Gamification , animated tutorial etc. The framework designed here is the outcome of the elearning experiences of the authors and it tries to add all relevant technologies with cutting-edge research to provide inspirational and transformative knowledge to learners of all ages, social status, communities who form worldwide communities of special-learners.

4. Sergio Serrano-Iglesias, ”Connecting formal and informal learning in Smart Learning Environments,”[4] 2021 - This article in a similar fashion to approaches related with mobile learning, the connection offered by SLEs can help students to reflect on learning concepts in real scenarios, but also adapting the offered resources to their progression and performance throughout the learning situation. However, existing attempts in SLEs face difficulties regarding the preparation of possible in- terventions by teachers or the understanding of the formal learning situation. This work attempts to overcome

5. Pitfalls of Machine Learning Methods in Smart Grids: A Legal Perspective [2021] International Symposium on Computer Science and Intelligent Controls (ISCSIC). The output of this paper is a visual

tool in form of a table, meant to guide prosumers, utility, technology and energy service providers. It shows the areas that need increased attention when dealing with specific prosumer concerns as identified in the technical literature.

6. A Computing Engine for the New Generation of Learning Environments [2020].IEEE 20th International Conference on Advanced Learning Technologies (ICALT). we proposed a computing framework, “Smart Learning Engine (SLEnG)

METHODOLOGY

The Pronunciation Checker system is designed to assist users in improving their speaking skills by leveraging deep learning techniques for real-time feedback and analysis. Upon logging into the platform, users are presented with randomly generated paragraphs that cover a broad spectrum of vocabulary, sentence structures, and linguistic nuances. This ensures that learners are exposed to a wide variety of words, enhancing their ability to handle diverse linguistic challenges.The system records the user's audio input and uses Automatic

Speech Recognition (ASR) to transcribe the spoken content. Once transcribed, deep learning models—such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs)—analyze the speech for any pronunciation errors. CNNs are particularly effective in capturing local patterns within speech data, while RNNs, with their ability to maintain context over sequences, help in understanding the temporal aspects of pronunciation. This combination of deep learning techniques ensures a highly accurate analysis of the user's pronunciation, detecting even subtle deviations from standard pronunciations.After the analysis, the system identifies mispronounced words and generates a comprehensive report that includes detailed phonetic corrections. The phonetic suggestions are provided alongside the meanings of the words, which allows users not only to correct their pronunciation but also to better understand the context and application of the words. This feature enhances both the learner's pronunciation and vocabulary development, making it a dual-purpose tool.The system also tracks the user's progress over time, compiling data into a personalized progress report. This report highlights areas of improvement and tracks overall pronunciation accuracy, providing users with a clear visual representation of their development. To enhance the learning experience, the platform offers percentage scores and progress metrics for each paragraph, allowing users to set benchmarks and continuously strive for improvement.A standout feature of the Pronunciation Checker is its adaptability to diverse accents and dialects. Through continuous updates to its deep learning models, the system becomes more attuned to various regional accents, offering tailored feedback that respects linguistic diversity. This feature ensures that users from different backgrounds can benefit equally from the tool, making it an inclusive solution for language learners worldwide.Furthermore, the user-friendly interface makes interaction seamless and intuitive, providing clear feedback without overwhelming the user with technical details. In addition to paragraphs, users have the option to record individual sentences for targeted feedback, focusing on specific aspects of their pronunciation that may need improvement. The real-time feedback fosters immediate learning, encouraging users to make corrections and practice repeatedly until the desired fluency is achieved.By incorporating continuous model training and updates, the Pronunciation Checker stays relevant to evolving language patterns and ensures that it remains a cutting-edge tool for enhancing speaking skills. Ultimately, the system is a comprehensive, adaptive platform that empowers users to gain confidence in their speaking abilities, making it an invaluable resource for language learners, professionals, and anyone looking to improve their pronunciation.

OBJECTIVE

The objective of the Pronunciation Checker system is to develop an intelligent, real-time tool that significantly enhances users' speaking and pronunciation skills using advanced deep learning techniques. This system is designed to accurately evaluate users' speech by analyzing theirpronunciation, identifying mispronounced words, and offering immediate, corrective feedback. It goes beyond simple error identification by providing users with phonetic transcriptions of the mispronounced words, detailed meanings, and context for each correction, which helps users not only fix their pronunciation but also expand their vocabulary and understanding of word usage. The system's key objective is to facilitate continuous progress tracking by maintaining a detailed history of each user's performance. This includes tracking the accuracy of their pronunciation over multiple sessions, generating performance reports, and visualizing improvement trends. By doing so, the system provides users with actionable insights into their language learning journey, helping them understand where they excel and which areas need improvement. Additionally, the Pronunciation Checker system is designed to be highly adaptable, with the ability to accommodate and adapt to various accents and speech patterns. This feature is crucial in ensuring that the system remains inclusive and effective for a diverse user base, offering personalized feedback that accounts for regional variations in speech while maintaining linguistic standards. Continuous model updates ensure that the system learns from user interactions and becomes more proficient over time in recognizing and assessing different accents, ensuring users receive accurate and tailored feedback regardless of their background. Another objective is to support effective language acquisition by creating a comprehensive and interactive platform that encourages consistent practice. Users can engage with the system through a variety of speaking tasks, such as reading randomly generated paragraphs or recording individual sentences for focused pronunciation practice. The real-time nature of the feedback encourages immediate corrections and repeated practice, fostering an environment conducive to active learning and rapid improvement. Moreover, the system is designed to be user-friendly, providing a simple yet effective interface that allows users to navigate and utilize the tool effortlessly. It offers features like audio recording, real-time analysis, and clear feedback presentation, ensuring that users at any level can benefit from the system without being overwhelmed by complexity. By blending deep learning models such as Convolutional Neural Networks (CNNs) for analyzing audio patterns and Recurrent Neural Networks (RNNs) for understanding temporal speech dynamics, the system delivers highly accurate and reliable pronunciation analysis. The end goal is to empower users to become more confident and proficient speakers by equipping them with the tools and insights necessary to master correct pronunciation, contributing to their overall language fluency and communication effectiveness.

PROBLEM DEFINATIONS

The problem addressed by the Pronunciation Checker system is the lack of accessible, real-time tools that effectively help language learners enhance their pronunciation. In traditional language learning environments, learners often rely on self-practice or occasional feedback from instructors, which lacks immediacy and specificity. This gap makes it challenging for learners to pinpoint exact mispronunciations or to recognize consistent errors in their speech. Without precise, real-time correction, many learners continue to make the same mistakes, hindering their progress and diminishing their confidence in spoken communication.

Furthermore, existing tools for pronunciation improvement typically provide only basic feedback and do not adapt well to the nuances of individual speech patterns or regional accents. These tools often operate with rigid models that cannot accommodate the wide diversity of pronunciations found in different dialects or language backgrounds, leading to frustration for users whose accents may not fit within the narrow

parameters of the system. As a result, learners from non-standard dialects or those speaking with regional variations may receive inconsistent and inaccurate feedback, limiting the effectiveness of such tools.

The Pronunciation Checker system addresses these challenges by leveraging advanced deep learning and speech recognition technologies to deliver accurate, real-time feedback tailored to each user's unique speech patterns. By analysing audio inputs with sophisticated models like Convolutional Neural Networks (CNNs) for sound feature extraction and Recurrent Neural Networks (RNNs) for temporal speech dynamics, the system can accurately detect and isolate specific pronunciation errors. This level of precision helps learners understand where they are going wrong, enabling them to make immediate corrections and avoid reinforcing incorrect habits.

Additionally, the system's adaptability to a wide range of accents and speech patterns ensures that it is effective for a global audience. It recognizes that language learners come from diverse linguistic backgrounds and provides feedback that respects these variations while guiding them towards standardized pronunciation. By accounting for different accents, the system offers personalized feedback that is both inclusive and accurate, making it a valuable tool for users worldwide

FLOW CHART

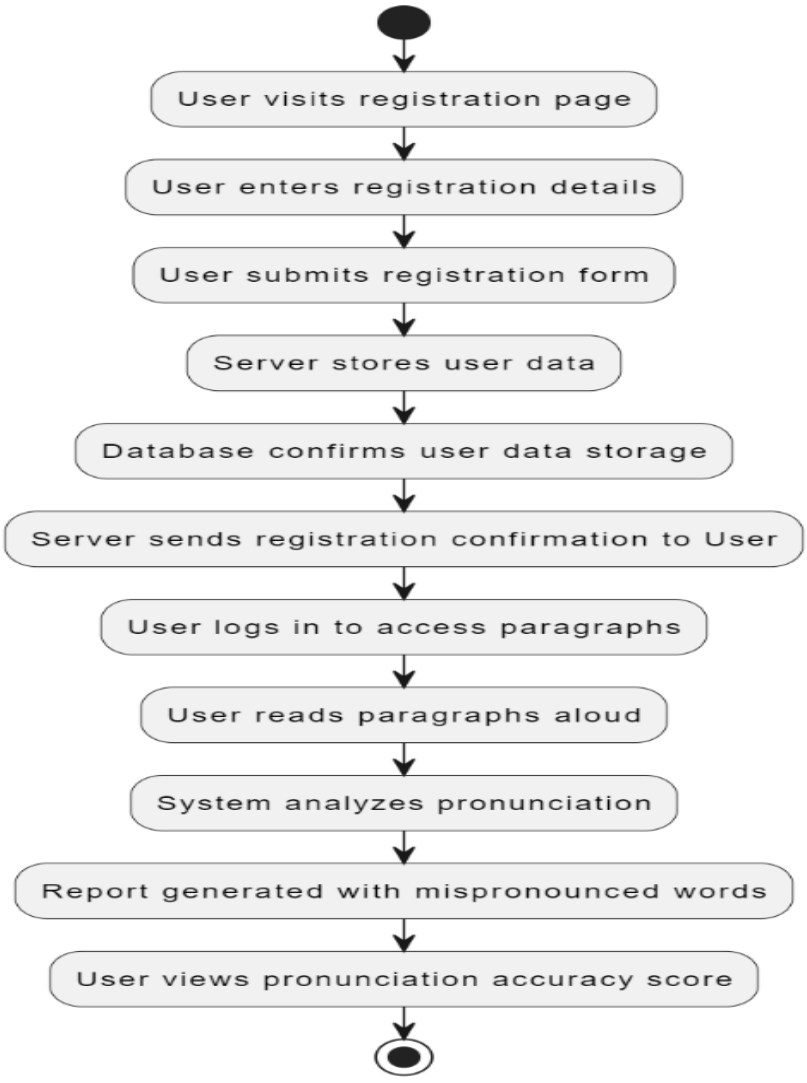
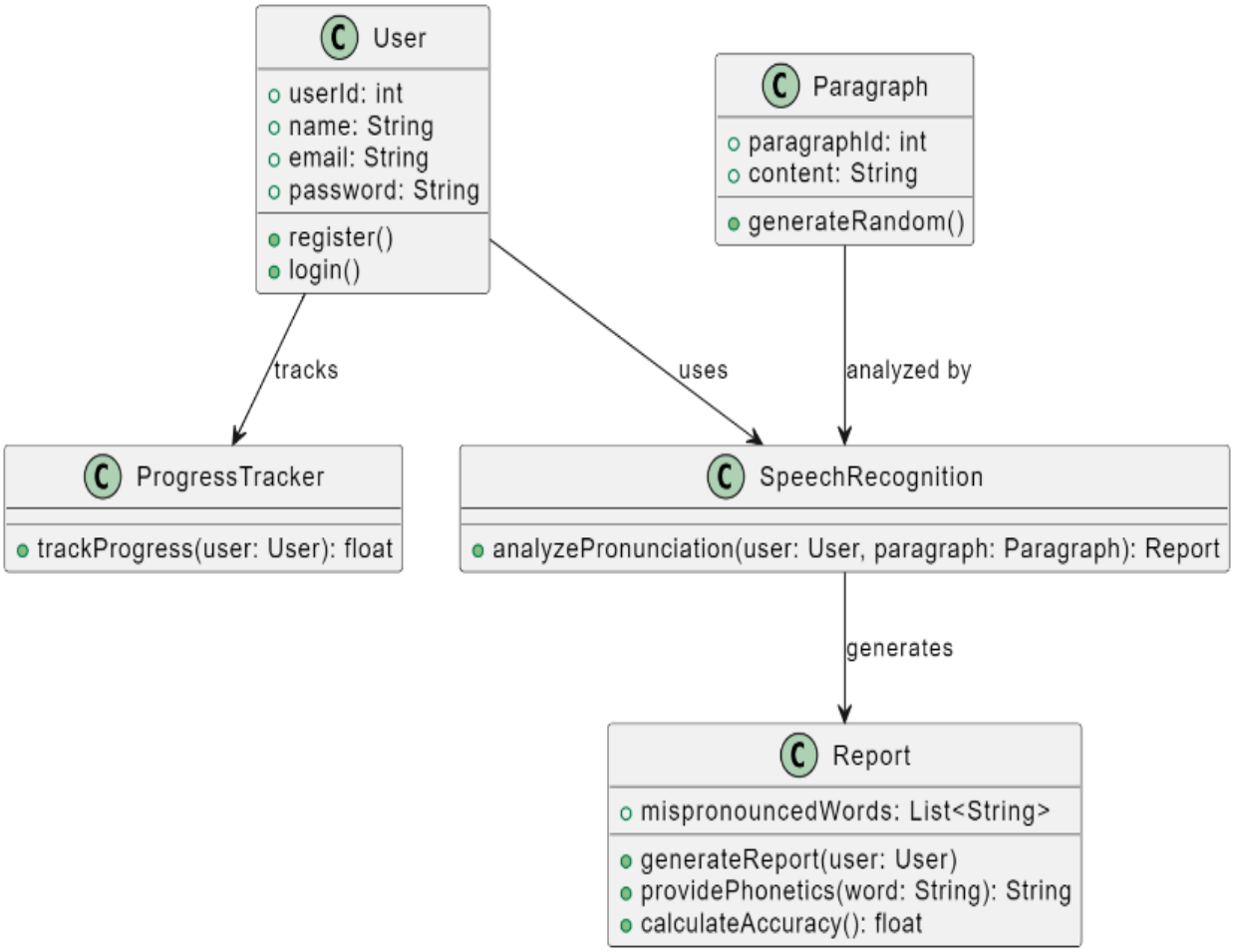


Fig 1: Flow Chart

MODEL DESIGNS

Uml Diagrams

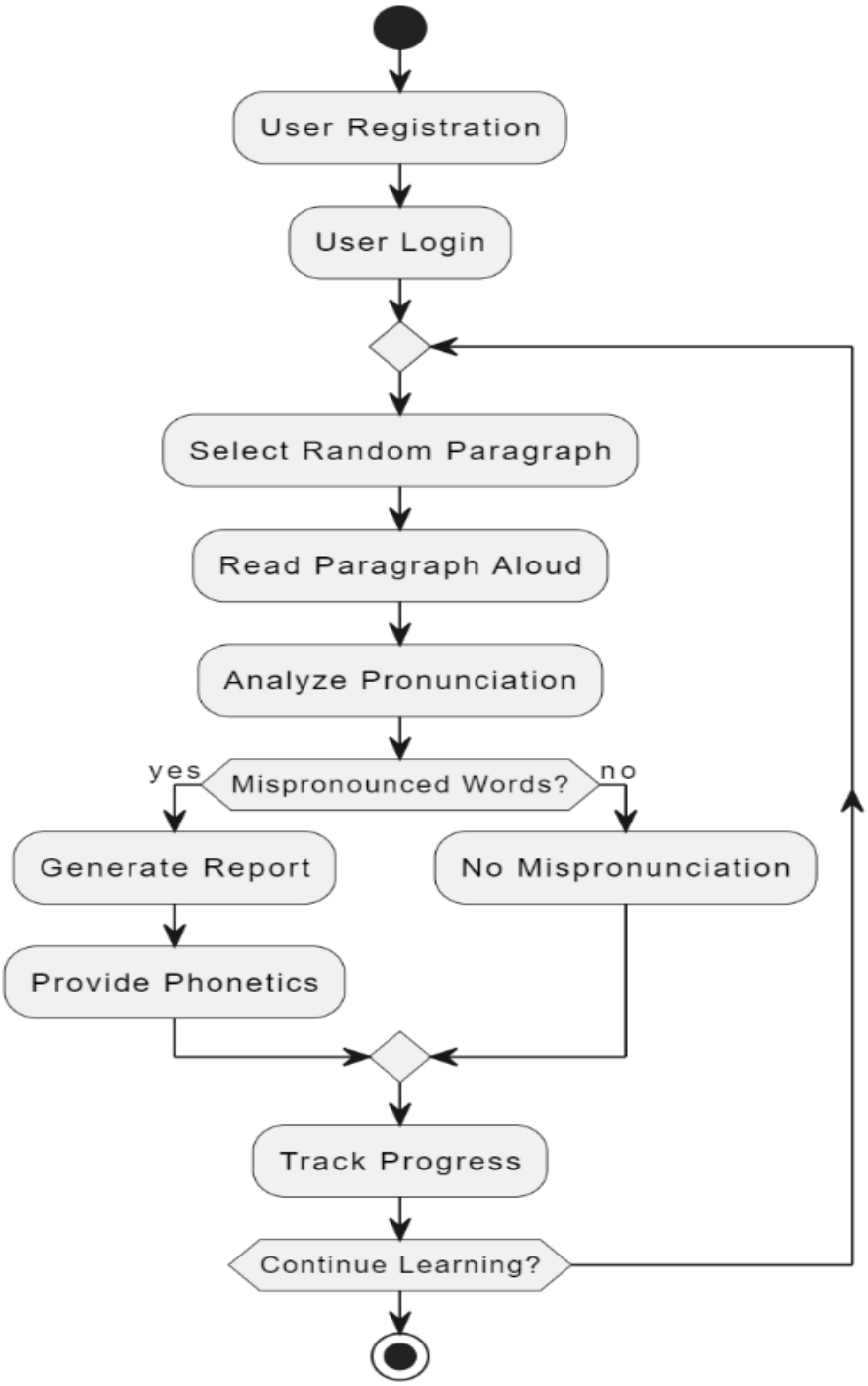
a) Class Diagram



Fig(2): Class Diagram

The class diagram represents a Pronunciation Checker system with five main components. The User class allows users to register, log in, and provides personal information. The Paragraph class generates random text for users to read, while the Speech Recognition class analyses the user's pronunciation using the provided paragraph. After analysis, it generates a Report that identifies mispronounced words, offers phonetic corrections, and calculates pronunciation accuracy. The Progress Tracker class monitors the user's progress over time, providing feedback on their improvement. This structure ensures an interactive and personalized experience for language learners.

b) Activity diagram



Fig(3): Activity Diagram

The activity diagrams outlines a process where a user registers and logs in, then reads a randomly selected paragraph aloud. The system analyzes their pronunciation for errors. If mispronunciations are found, a report is generated with phonetic guidance. Progress is tracked, and the user can choose to continue learning by repeating the process or stop if no further learning is desired.

c) Sequence diagram

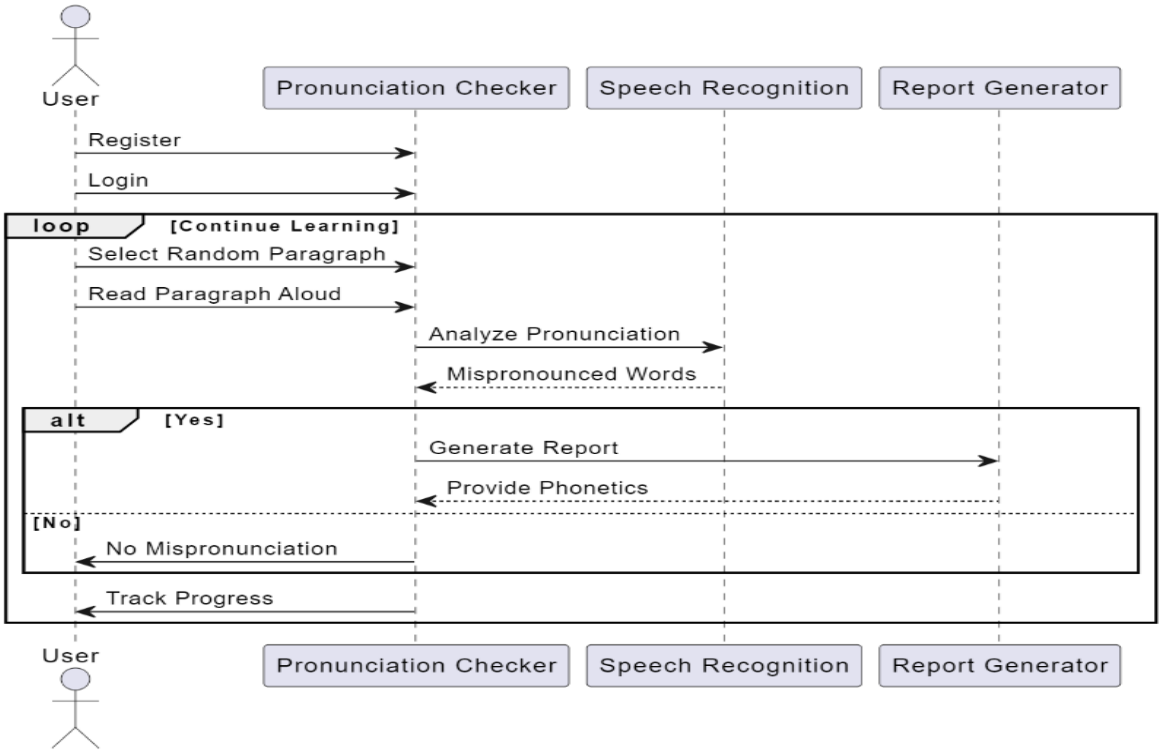


Fig (4): Sequence Diagram

The sequence diagram shows how a user registers, logs in, reads a random paragraph aloud, and has their pronunciation analyzed. If mispronunciations are found, a report with phonetic feedback is generated. If there are no errors, progress is tracked. The process loops until the user decides to stop. The interaction involves three systems: Pronunciation Checker, Speech Recognition, and Report Generator.

d) Use case diagram

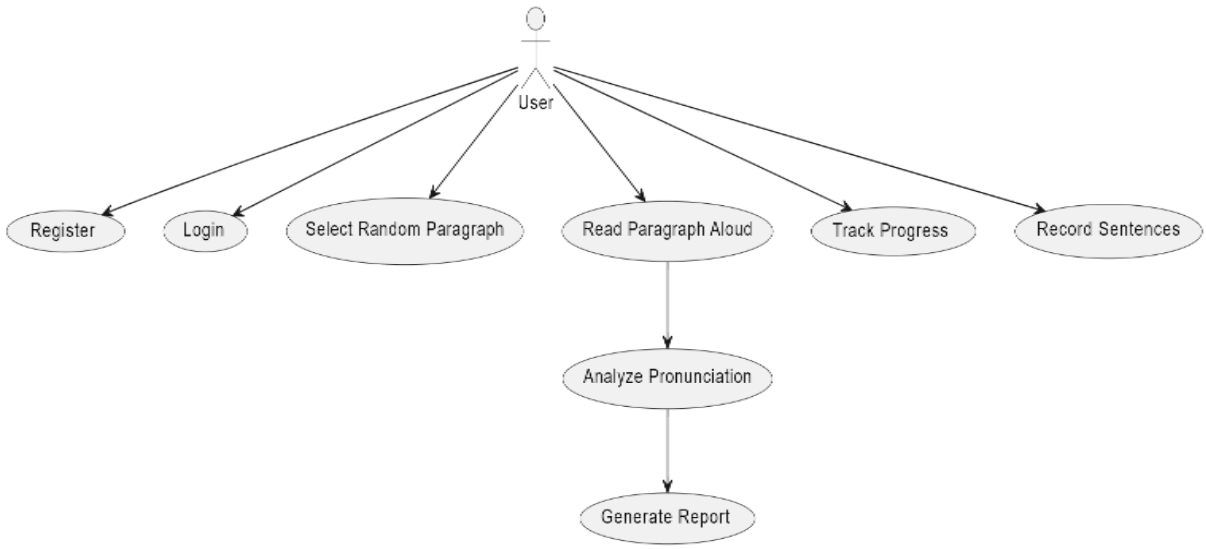


Fig (5): Use Case Diagram

The use case diagram illustrates a user performing key actions such as registering, logging in, selecting a random paragraph, and reading it aloud. The system then analyzes the user's pronunciation and generates a

report based on any detected errors. Additionally, the system tracks the user's progress over time and allows them to record sentences for practice.

These interactions help the user improve their pronunciation, with real-time feedback and progress monitoring to guide their learning journey.

e) Package diagram

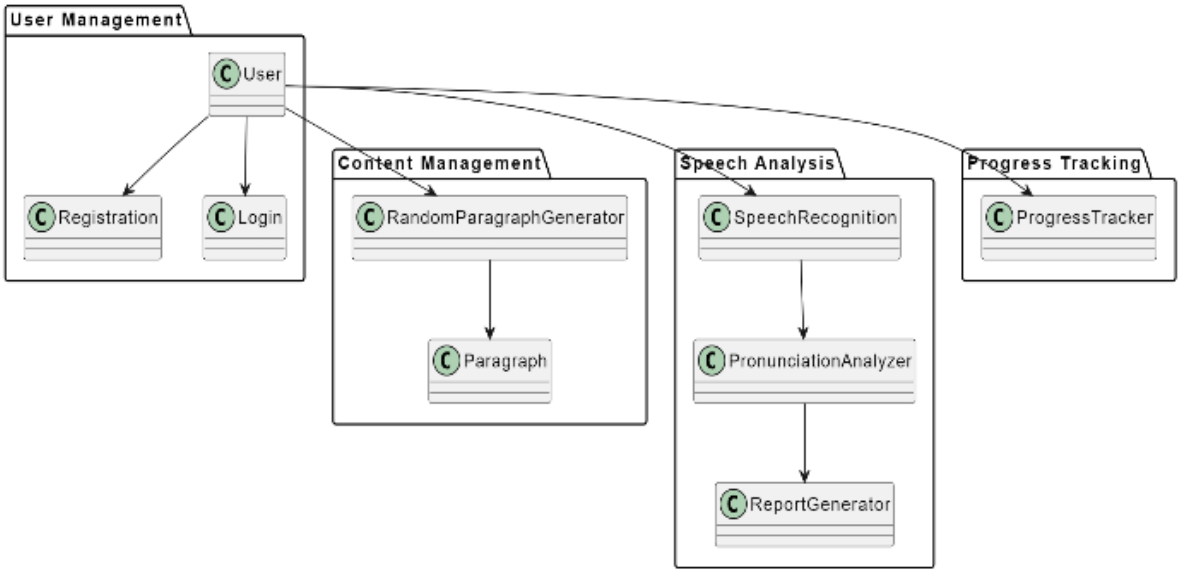


Fig (6): Package Diagram

The diagram illustrates a system with four key components: User Management, Content Management, Speech Analysis, and Progress Tracking. User Management handles registration and login via the User class. Content Management generates random paragraphs through the RandomParagraphGenerator and Paragraph classes. Speech Analysis involves speech recognition and pronunciation analysis, producing reports. Progress Tracking monitors user performance through a ProgressTracker component.

FUCTIONAL REQUIREMENTS

- The system must allow users to register with their personal details such as name, email, and password, and must ensure secure authentication and authorization during login with valid credentials to access the pronunciation checker. It should include features like password recovery and user profile management for a seamless experience.
- The system must generate random paragraphs from a curated dataset that includes sentences of varying complexity and phonetic challenges. These paragraphs should be customizable based on user level (beginner, intermediate, advanced) to provide appropriate difficulty for improving pronunciation skills.
- The system must be able to record the user’s speech in real-time using a web or mobile interface, ensuring that the interface is intuitive and responsive across devices. It should capture high-quality audio, and include noise cancellation and volume control features to ensure the clarity of the recorded speech.
- The system must transcribe spoken words using Automatic Speech Recognition (ASR) models, converting the recorded audio into text. The ASR models should be capable of handling diverse accents and variations in pronunciation to ensure accurate transcription. It should also support various languages and dialects to broaden accessibility.

- The system must compare the transcribed speech with correct phonetic representations stored in a phonetic dictionary or database, using deep learning models such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) to detect mispronunciations. It should provide high accuracy in identifying subtle pronunciation errors.
- The system must generate a detailed report for each session, highlighting the words that were mispronounced and providing phonetic corrections along with audio feedback for correct pronunciation. The report should include word meanings, contextual usage, and personalized pronunciation tips. Additionally, the system should track user progress over time, presenting improvements and areas that still need attention.
- The system must support user progress tracking by storing historical data of previous sessions, enabling users to view their improvement in pronunciation over time. This feature should include visual aids like graphs or progress bars to motivate continued learning.
- The system should allow users to repeat specific exercises or focus on words or sounds they find challenging, offering customizable learning paths. This ensures personalized learning experiences based on individual needs.
- The system must provide users with helpful resources, such as video tutorials or tips on correct pronunciation techniques, which could be accessed alongside their feedback reports for continued improvement. These resources should be available through the system's user interface for easy reference.

NON FUCTIONAL REQUIREMENTS

- The system must ensure high accuracy in recognizing spoken words by utilizing advanced speech recognition algorithms capable of handling various accents and speech patterns. The deep learning models should be continuously trained on diverse datasets to improve recognition accuracy, ensuring that the system provides reliable feedback on pronunciation regardless of regional or linguistic variations.
- The system must provide real-time or near-real-time feedback to users immediately after they finish reading a paragraph or recording a sentence. This feedback should be processed quickly, ensuring an interactive and engaging experience where users can immediately see their results and make corrections without delays, enhancing the learning process.
- The system must be designed to handle multiple users simultaneously, ensuring that the performance remains stable and efficient, even under high traffic. This includes optimizing server performance, load balancing, and ensuring the underlying architecture can scale to meet demand without affecting the quality of the feedback or analysis.
- The system must ensure secure storage and transmission of user data, including audio recordings, progress reports, and personal information. Data encryption (both in transit and at rest) should be implemented, along with compliance with relevant data protection regulations (such as GDPR or CCPA) to safeguard user privacy and ensure the system maintains high standards of security.
- The system must be adaptable to various accents and dialects, ensuring that it delivers relevant and personalized feedback for users from diverse linguistic backgrounds. The system should leverage a robust and diverse dataset that reflects global linguistic variations, and the speech recognition models should be fine-tuned to minimize bias toward any particular accent or speech pattern.
- The user interface (UI) must be intuitive and user-friendly, designed with accessibility in mind. This includes clear navigation, minimal learning curve, and visual aids to guide users through the pronunciation exercises. The UI should cater to users with varying levels of technical expertise, ensuring that even non-tech-savvy individuals can easily engage with the system.

- The system must be easy to update and maintain, with a modular design that allows for the retraining and improvement of machine learning models over time. This includes a clear update process for the speech recognition and deep learning models, ensuring that the system stays up-to-date with the latest advancements in AI and speech technology.
- The system must incorporate feedback mechanisms where users can provide input on their experience and report any challenges with pronunciation detection. This user feedback can be used to continuously improve system performance, ensuring that it remains effective and relevant to a diverse range of users.
- The system must support multilingual capabilities, allowing users to practice pronunciation in multiple languages. This feature would broaden the system's appeal, making it useful not only for English learners but for individuals working on other languages, ensuring a global reach.
- The system must be built with cross-platform compatibility, functioning seamlessly across different devices such as desktops, tablets, and smartphones. This ensures users can access the pronunciation checker on the go, making learning more flexible and convenient.

CONCLUSION

In conclusion, the Pronunciation Checker system offers an innovative and comprehensive solution to the challenges faced by language learners in improving their speaking and pronunciation skills. By integrating advanced deep learning techniques with cutting-edge speech recognition technology, the system provides real-time, accurate, and personalized feedback on pronunciation, tailored to individual learning needs. It identifies specific mispronunciations, delivers precise phonetic corrections, and offers detailed explanations of word meanings, enabling users to understand and rectify their mistakes effectively.The system’s ability to track user progress over time further enhances its value, fostering consistent improvement and ensuring learners remain motivated throughout their language learning journey. Its adaptability to various accents and dialects ensures inclusivity, providing relevant feedback to users from diverse linguistic backgrounds. The intuitive, user-friendly interface makes the tool accessible to users of all ages and technical proficiencies, enhancing its usability and engagement.Furthermore, the comprehensive progress reports and feedback system enable learners to identify patterns in their pronunciation errors, allowing for targeted practice and improvement. The system’s cross-platform functionality and secure data management make it a reliable and flexible tool that can be used on different devices and in various settings, catering to the needs of modern learners.Overall, the Pronunciation Checker system not only enhances language acquisition but also builds confidence in spoken communication, making it an invaluable resource for students, professionals, and anyone looking to improve their pronunciation and fluency in any language.

REFERENCES

1. Pitfalls of Machine Learning Methods in Smart Grids: A Legal Perspective, Gorge jems, cleadecker ,cloe, 2021
2. A Computing Engine for the New Generation of Learning Environments, Alfred joe, Taylor mirinda,, 2020
3. Developing a Virtual Smart Total Learning Environment for Future Teaching-Learning System, Allena ,trexishaen, 2020
4. Connecting formal and informal learning in Smart Learning Environments, Seeminniclors ,rafal john , 2021
5. Pronunciation Scoring With Goodness of Pronunciation and Dynamic Time Warping, KavitaSheoran; ArpitBajgoti; Rishik Gupta; Nishtha Jatana; Geetika Dhand; Charu Gupta, 13 February 2023

6. Pronunciation Modeling With Reduced Confusion for Mandarin Chinese Using a Three-Stage Framework, Ming-Yi Tsai; Fu-Chiang Chou; Lin-Shan Lee, 22 January 2003
7. Supervised Detection and Unsupervised Discovery of Pronunciation Error Patterns for Computer-Assisted Language Learning , Yow-Bang Wang; Lin-shan Lee, 05 January 2015
8. Japanese Pronunciation Evaluation Based on DDNN, Deguo Mu; Wei Sun; Guoliang Xu; Wei Li 2 December 2020
9. Stochastic Pronunciation Modeling for Out-of-Vocabulary Spoken Term Detection
10. State-dependent phonetic tied mixtures with pronunciation modeling for spontaneous speech recognition , Yi Liu; P. Fung, 31 july 2004