Raabta

Kaavish Project Proposal

By

Fatima Tariq	(ft07200@st.habib.edu.pk)
Muminah Khurram	(mk07521@st.habib.edu.pk)
Raza Hashim	(rn07380@st.habib.edu.pk)

September 29, 2024

In partial fulfillment of the requirement for

Bachelor of Science

Computer Science

Dhanani School of Science and Engineering

Habib University

Fall 2024

1 Abstract

In our society, communication is a fundamental pillar of interaction and connection, yet millions of individuals who are deaf or hard of hearing face substantial barriers to fully integrating into their communities. In Pakistan alone, approximately 10.2 million individuals are deaf or hard of hearing and would rely on Pakistani Sign Language (PSL) to communicate, yet the broader population remains largely unversed in PSL, rendering communication with the deaf community incredibly challenging [1]. The existing solutions, such as interpreters or technology like sign language translation gloves, are both financially prohibitive and socially alienating. These tools, while well-intentioned, further isolate the deaf by placing the burden of bridging communication gaps squarely on their shoulders. Consequently, the responsibility of inclusion is placed disproportionately upon deaf individuals, leaving them excluded from mainstream societal participation and exacerbating the communication divide.

Moreover, educational opportunities for deaf individuals are notably constrained. The limited availability of resources for learning PSL, alongside the passive nature of existing learning tools—often based on video demonstrations alone—do not provide the interactive experience necessary for effective language acquisition. It is widely acknowledged that fluency in any language, including sign language, cannot be achieved solely through observation; active practice and real-time feedback are crucial for this. Unfortunately, the cost of PSL classes and their inflexible schedules present additional obstacles for those who wish to learn.

The individuals most affected by these limitations are not only those who are deaf but also their families, educators, and broader society, all of whom would benefit from more seamless and inclusive communication. Despite the significant number of people who could stand to gain from learning PSL, the scarcity of the right learning tools prevent many from pursuing it. As a result, deaf individuals often find themselves excluded from educational and career opportunities, while those who wish to interact with them are unable to do so in a meaningful way.

The proposed objective of our project, Raabta, is to address these challenges by offering an innovative solution that redefines how PSL can be learned and practiced by incorporating real-time computer vision technology to provide live interactive feedback, enabling users to engage with PSL in a more practical and dynamic manner. The app's emphasis on active participation allows learners to correct their signing mistakes instantaneously, improving both accuracy and fluency. Accessible through smartphones, Raabta democratizes the process of learning PSL, making it possible for a wider audience to learn at their own pace and convenience, without the financial and logistical barriers that currently limit PSL education and by making PSL more accessible to all, Raabta has the potential to create a more inclusive society, wherein communication is no longer a privilege reserved for the hearing but a right shared by all.

2 Problem definition

2.1 Background and Current Challenges:

The ability to communicate using sign language is crucial for the social inclusion and empowerment of the deaf and hard-of-hearing community. According to an estimation by the World Health Organization Pakistan has a population of approximately 10.2 million hearing-impaired people, of whom only a small fraction of less than 5%, receive formal education, contributing to their social isolation [1, 2]. The lack of access to sign language learning opportunities exacerbates this problem,

particularly in an educational and social landscape that is not conducive to deaf inclusion. While Pakistani Sign Language (PSL) is the most widely used sign language in the country, learning it, like many sign languages globally, presents several unique challenges.

One of the most significant barriers to learning sign language is the misconception that it is a simplified, picture-like form of communication. In reality, sign language is as complex and nuanced as any spoken language, involving intricate hand movements, facial expressions, and body language. Developing an expertise in these elements requires not only an understanding of vocabulary but also extensive practice to develop fluency. Existing resources for learning PSL are limited and often inadequate with most platforms offering only sign language dictionaries or video-based tutorials but lack interactive features or opportunities for practice. This one-sided mode of learning is insufficient for developing the practical skills required for fluent communication in sign language.

Additionally, due to the social stigma and the lack of educational infrastructure for the deaf, there are few spaces where learners—both deaf and hearing individuals—can regularly practice PSL and receive feedback on their progress. This creates a learning environment that is often isolating and ineffective, ultimately discouraging individuals from pursuing PSL mastery.

2.2 Proposed Solution:

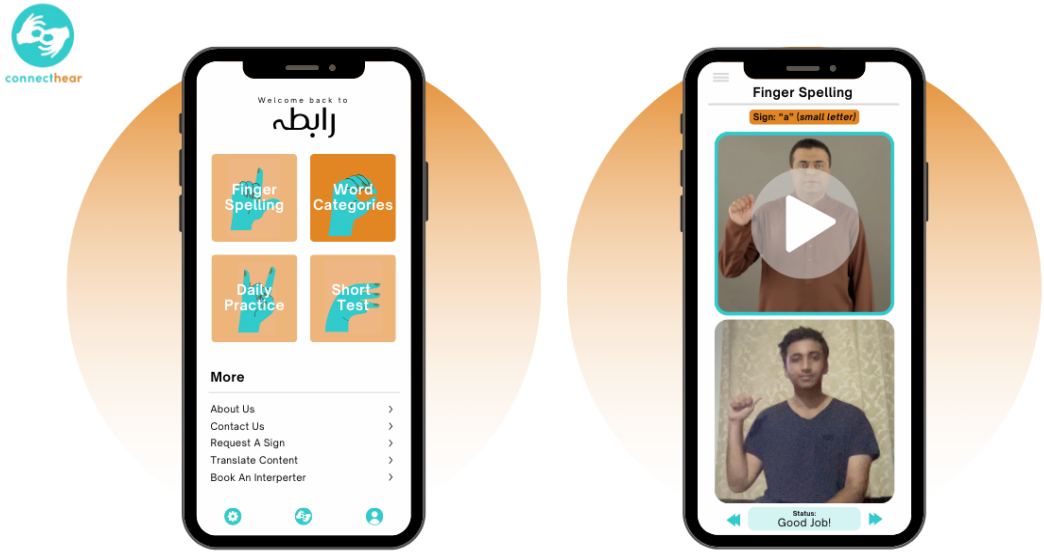

Figure 1: Proposed Mobile Application Wireframes

To address these challenges, we propose the development of a mobile application, **Raabta**, designed to facilitate the learning and practice of Pakistani Sign Language in an interactive, engaging, and accessible manner. Our solution is inspired by popular language-learning applications, such as Duolingo, but tailored specifically for PSL learners. Raabta will be developed in collaboration with **ConnectHear**, a social enterprise committed to addressing the issues faced by the deaf community,

whose deep understanding of the community's needs will guide the project's direction and ensure the platform's inclusivity.

Raabta will offer structured lessons aimed at teaching PSL, but it will also address the crucial need for practice by allowing users to interact with the app using real-time signing. The core innovation of this application of real-time feedback through advanced computer vision and machine learning techniques will allow users to practice signing by performing signs in front of their phone's camera, and the application will analyze their movements, facial expressions, and body language to provide instant feedback on the correctness of their signs. This will significantly enhance the learning experience by simulating real-world practice, something that all existing platforms fail to offer. Moreover, the app will be designed with a focus on accessibility, ensuring that learners from diverse backgrounds, including those with limited technological experience, can easily use and benefit from it.

By creating this platform, in partnership with ConnectHear, we aim to reduce the barriers to learning PSL and improve the overall inclusivity of Pakistan's deaf population by making the language more accessible to both deaf and hearing people in Pakistan. By providing a convenient platform for learning PSL, we aim to promote deaf-inclusion and empower individuals to communicate effectively, thus contributing to the broader societal goal of reducing the marginalization of the deaf community in Pakistan.

3 Social relevance

- **Improves Self-Esteem and Empowerment:** Accessible PSL learning platforms boost the self-esteem of deaf individuals by enabling them to communicate more freely. ConnectHear's efforts to provide educational opportunities for the deaf community have already laid the groundwork for this empowerment, and our platform will build on this by offering interactive learning tools that can reach even more people.
- **Shifts Social Responsibility:** The project helps shift the burden of overcoming communication barriers and social stigma from the deaf community to the hearing population. By encouraging more people to learn PSL, it promotes mutual understanding and reduces societal biases against the differently abled.
- **Promotes Social Inclusion:** Learning PSL raises awareness about the challenges faced by deaf individuals in daily life, promoting greater tolerance, acceptance, and diversity in society. Additionally, by breaking down communication barriers between hearing and deaf individuals, we aim to create a more open and accepting society where deaf or hard-of-hearing individuals can engage freely without feeling marginalized.
- **Reduces Educational and Employment Inequality:** The app addresses the educational barriers faced by those trying to learn sign language, offering them opportunities to learn PSL and improve their literacy. People with communication disabilities are often economically disadvantaged, and by making sign language learning more accessible, the project helps reduce some of the economic barriers tied to communication difficulties, leading to improved job prospects and societal integration.
- **Cognitive and Educational Benefits:** Learning sign language has been shown to enhance cognitive abilities such as memory, problem-solving, and spatial reasoning. For hearing individuals, becoming bilingual in both spoken and signed languages fosters analytical thinking,

multitasking skills, and improved social cognition, benefiting overall personal development. In the long run, it can even protect against cognitive decline in old age, including delaying the onset of dementia by 4 to 5 years [3].

4 Originality/Novelty

Raabta addresses the critical issue of effective communication with the deaf and hard-of-hearing community in Pakistan. By utilizing advanced computer vision and machine learning/deep learning technologies for real-time feedback, this app offers an innovative and interactive learning experience that is currently lacking in existing solutions.

Existing sign language learning platforms typically rely on video tutorials or dictionaries, which lack interactivity and do not allow learners to practice and receive feedback on their signing skills. For instance, platforms like "PSL by Deaf Reach" or even free mobile apps for learning other variants of sign language like BSL or ASL provide video resources but do not offer real-time feedback mechanisms, making it challenging for users to develop fluency [4].

In contrast, SignTutor, while offering a structured approach to learning sign language, requires users to wear colored gloves for accurate gesture recognition [5]. It works using a vision-based approach. This reliance on an external resource can limit usability and user experience, as it imposes additional costs and inconvenience. Furthermore, SignTutor's feedback primarily focuses on hand gestures and movements, overlooking the significance of facial expressions and body language, which are integral components of sign language communication.

Our application distinguishes itself by eliminating the need for gloves or any such external resources allowing users to practice signing using only their hands, body and face. By analyzing movements and expressions in real-time solely through their device's camera, making it far more accessible than existing solutions. The app will also provide feedback on both manual and non-manual components of signing to aid in the learning process.

Upon reviewing the current research, it was noted that many existing sign language recognition systems are sensor-based, often relying on data gloves or other wearable devices for accurate gesture detection. While these systems have shown some success, they are inherently restrictive and impractical for large-scale, everyday use due to their dependency on external sensors. A review by Zahid et al. highlighted that machine learning and vision-based systems, particularly for Urdu and Pakistani Sign Language, are still underdeveloped, with limited publicly available datasets for PSL [6]. Moreover, approaches that rely on external sensor-based hardware introduce an additional significant cost that inhibits large portions of the community. Such hardware also further adds to the social stigma as it adds the burden of carrying around specialized equipment to the already marginalized community further highlighting their difference. Instead of this, we believe efforts should be focused on encouraging the hearing person to make the effort to learn sign language to make deaf people feel more included.

In addition to the existing solutions, a notable study on PSL recognition proposed a pipeline utilizing deep learning models like C3D, I3D, and Temporal Shift Module (TSM), demonstrating that data augmentation techniques to combat the limited dataset size such as rotation and translation improved model performance, with C3D achieving an accuracy of 93.33% [7]. This study also used the PSL dictionary dataset, which contains signs for over 6,000 words; however, only 80 signed words from the top 1,500 most frequently used words in Urdu were selected, limiting the model's scope. While this study is a major one in applying advanced action recognition models to

PSL, it focused primarily on sign recognition rather than an interactive learning experience as well as being highly limited in terms of data used.

5 CS contribution

- The solution we are proposing will be in the form of a mobile application. Our choice of a mobile application is twofold: first, it stems from our desire to make it accessible to as much of the population as possible, and secondly, there are more people with smartphones than personal computers. Due to this, we will be requiring knowledge and skills related to mobile app development. Courses such as **Web and Mobile Development** will be crucial for executing this part of the project along with **Software Engineering**.
- The system we are proposing will have to pick up the gestures that are being made through an individual's hands, body, and face. To do so, we will have to make use of Computer Vision for the system to be able to identify the relevant body parts and track their motion to detect the gestures they are making. Currently, our group has a rich understanding of **Deep Learning**, but will be making additional efforts to better our experience with Computer Vision through self-learning.
- For identifying whether the gestures made are correct we will have to rely on Deep Learning or Machine Learning frameworks. The gestures will have to be matched with their closest recognized gesture. This will lead us down two paths, include making use of machine learning approaches to cluster the vectorised data or using traditional Deep Learning models
- The general working of the project would also require us to work with a lot of data due to our machine-learning-based approach. This would require extensive pre-processing of data to make it relevant and useful to us. This would mean that our experience with courses such as **NLP** and **Data Science** would come in very handy here.
- Our group's extensive experience with related areas such as Large Language Models (LLMs), Natural Language Processing (NLP) and Deep Learning we believe will make it easier for us to utilize these machine-learning approaches as well. Our experience is further augmented by our knowledge in the field of Graphical Data Science which we may employ for the classification of signs.
- The extensive amount of data attached to our project would also require us to make use of our skills learned in the **Database Management System** course. This would allow us to store all the PSL-related text and their associated sign gestures effectively and access them efficiently when needed.
- For the success of the project, maintaining an efficient workflow and collaborative environment would be crucial factors. Not only would we need to manage internal group dynamics effectively, but the involvement of an external entity in our project would require us to understand and manage their needs too. Making the skills we have picked up from courses such as **Software Engineering** and **Engineering Project Management** would be vital here.

6 Scope and Deliverables

The problem we aim to solve is a highly complex one with little to no work done on it, especially in the local context. This would mean that we would have to make use of the limited data we have to ensure the accuracy of our systems. This would require data scraping to build our database as well as extensive pre-processing of the data gathered and create mappings between instructional videos to PSL vocabulary. The recognition of the user's gestures would require the deployment of computer vision libraries to pick up their gestures in real time. The comparison of the user's gestures and the recognized vocabulary would require the usage of Machine Learning Algorithms as well as data augmentation to account for differences in placement, position, orientation, lighting, and addition of noise to improve overall reading accuracy on a better-trained model. We believe that scope is incredibly expansive and would take a significant amount of time to get through while keeping in mind our communication and requirement acquisition with ConnectHear. Besides this, we would also have to develop a user-friendly app that will have curated lesson plans as well as practice and appropriate testing to ensure user certification is granted to qualified users. The incorporation of all these features and having an app that has the potential to be deployed would require time spent on testing the system and application with our industry partner to ensure that it matches the needs of the users as they have a better understanding of that than we do. We believe all of this is a big undertaking that has the potential to have great social significance too, and we hope that it can lead to the deployment of a learning application under the guidance of our external partner, ConnectHear.

Deliverables:

- **Mobile Application:** The final product of our project would be in shape of a mobile application we will develop. This user-friendly application will aim to make sign language learning accessible to a large chunk of the population. We will try and make the interface as accessible as possible. We aim to ensure that the application has the functionality to add newer vocabulary with time along with improved lessons.
- **Database:** For our data we will refer to the psl.org.pk's existing digital library where we will scrape relevant data to create a database to organize PSL vocabulary and corresponding signs. This database would additionally serve as a very helpful starting point for future work in the field [8].
- **Translation Model:** The application would rely on a mechanism to match the users' real-time gestures to a pre-existing repository of recognized PSL signs. This model is greatly novel in the current landscape and can be improved upon down the line for better feedback, improved accuracy, and recognition of more signs.
- **Mapping System:** To support our application we will have to develop a backend mapping of sign language videos to PSL vocabulary. This mapping can prove useful for further work in the field. To identify whether the gestures that the users are making are correct or not we will have to formalize the distribution of examples of proper and recognized signs in PSL which the users' gestures can be compared to. This distribution can serve as a foundation which can be augmented by adding more examples and more signs down the line to expand its usability and allow it to be used for other similar projects down the line.

- **Documentation:** Along with the application, we will also have to develop extensive documentation to make it accessible for the users but also documentation that would allow other entities to expand its scope and make it accessible to greater chunks of the population.

Feasibility

Dataset

The primary dataset for this project will be obtained by web-scraping PSL (Pakistan Sign Language) data from psl.org.pk. This website contains over 6,000 PSL signs, which will serve as the foundational dataset for training machine learning models.

Data Acquisition

The web scraping process will involve extracting relevant PSL signs, corresponding videos, and metadata using custom scraping scripts. Additional and more specific PSL data will be provided by ConnectHear to supplement the scraped dataset.

Data Preprocessing and Augmentation

The raw data obtained from the website is a video of the signer practicing a sign twice with primary use of their right hand on a clear background in clear quality. This data will be augmented by splitting the two signs in a single video into two separate videos, flipping the videos, scaling, rotating, adding noise, changing colors, and further data augmentation techniques. This dataset will be cleaned and used for training the machine learning model.

Compute Resources

Cloud Computing Platforms

- **Google Colab**: Used for initial model development and training; we might need Colab Pro depending on the model and dataset size.
- **Amazon Web Services (AWS)**: Utilized for larger-scale training and deployment of machine learning models, providing scalable resources.

Data Storage Solutions

- **Cloud Storage**: Datasets and trained models will be stored in AWS S3 or Google Drive for easy access and scalability.

Software Libraries

- **Machine Learning and Computer Vision**: TensorFlow or PyTorch will be used for model development, while MediaPipe Holistic and OpenCV will be employed for gesture recognition and video processing.

- **Mobile App Development**: React Native or Flutter will be used for developing the mobile application, along with Firebase for managing the backend.

Hardware

- **Standard Smartphones**: As the app will be targeted at mobile devices, no special hardware is required beyond a mid-range smartphone with a camera, which will be the primary device for using the app.
- **Development Laptops/PCs**: Regular computers will be sufficient for app development, while cloud-based resources will be used for heavier computations (such as model training).

7 Team dynamics

- **Fatima Tariq:** Due to her experience with the course of web and mobile development, she serves as the most obvious choice to lead the development side of things for our final product. Her experience in leading the software engineering project further strengthens our belief in our choice. Due to her previous success in curating an important and effective dataset for a Deep-Learning project and her experience of having taken the Data Science course at the University of California, Berkeley, she is the most qualified person in our group to deal with the large amounts of data that we will have to handle. She will curate the dataset that we will require by scraping the data, augmenting it, and organizing it.
- **Muminah Khurram:** Research is the key to success for any Final Year Project, that is why we will have a team member specifically in charge of that and due to her vast experience with it, she serves as the logical choice. She will also be the person responsible for the front-end of the project with focus on making the platform intuitive and accessible for end-users. Besides that, due to having taken courses such as Social Network Analysis and Graph Data Science, she will also be the primary person responsible for the implementation of the Machine Learning aspect of the project including exploration of deep learning models, data vectorization, mapping and clustering.
- **Raza Hashim:** On the technical side of things, due to his past experience with Deep-Learning models he will serve as the person responsible for the computer vision pipeline. Besides the technical side, due to his wealth of managerial experience in various contexts, he will serve as the project manager and oversee the entire operation, and this choice is further augmented by the fact that he did well in the Project Management course. His extracurricular experience has also led to him dealing with people from varying backgrounds, which also makes him the logical choice to serve as the liaison with the external partner.

Tech Stack

Mobile Application Development

Framework

- **Flutter**: For building cross-platform mobile apps (Android & iOS) with a single codebase.

- **Dart**: The primary programming language used with Flutter.
- **Django**: For backend services.

UI/UX

- **Dart UI**: For creating a responsive and user-friendly interface in Flutter.
- **Flutter Widgets**: For building rich user interfaces with customized components.

Machine Learning and Computer Vision

Frameworks

- **TensorFlow Lite**: For on-device machine learning inference to handle gesture and facial recognition.
- **MediaPipe Holistic**: Integrated with TensorFlow Lite for real-time body, hand, and face tracking.

Model Training & Preprocessing

- **OpenCV**: For image preprocessing tasks such as augmentation and gesture detection.
- **Google Colab or AWS EC2**: For larger-scale model training.
- **Pytorch/TensorFLow**: For implementing machine learning algorithms/deep learning models for gesture matching.

Database

- **Firebase Real-Time Database**: For real-time user data and learning progress.

Cloud Storage & Deployment

- **Google Cloud Storage**: For storing assets, such as videos and machine learning models.
- **Firebase Cloud Functions**: For handling serverless functions, notifications, and request processing.
- **Amazon S3**: For storing PSL videos, images, and datasets.
- **Heroku**: For deployment.

Data Scraping and Preprocessing

- **BeautifulSoup or Scrapy (Python)**: For scraping PSL videos and metadata from psl.org.pk.
- **Pandas**: For managing and cleaning the data collected, preprocessing the metadata, and organizing it into structured formats.
- **openCV**: For augmenting video data by flipping, rotating, scaling, adding noise, and other transformations

References

[1] AFP, "For Deaf Children in Pakistan, School Is Life," DAWN.COM, 29 May 2024. Available: www.dawn.com/news/1836327. [Accessed: 06-Oct-2024].

[2] World Health Organization, "Deafness and Hearing Loss." Available: https://www.who.int/health-topics/hearing-loss.

[3] H. Bowman-Smart et al., "The Moral Case for Sign Language Education," Monash Bioethics Review, vol. 37, no. 3-4, pp. 94–110, 2019. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6900220/, doi: https://doi.org/10.1007/s40592-019-00101-0.

[4] PSL by Deaf Reach, Pakistan Sign Language, Dec. 29, 2022. Available: https://apps.apple.com/us/app/psl-by-deaf-reach/id1619629318. [Accessed: 06-Oct-2024].

[5] O. Aran et al., "SignTutor: An Interactive System for Sign Language Tutoring," IEEE MultiMedia, vol. 16, no. 1, pp. 81–93, Jan. 2009. Available: https://doi.org/10.1109/mmul.2009.17.

[6] H. Zahid, M. Rashid, S. Hussain, F. Azim, S. A. Syed, and A. Saad, "Recognition of Urdu sign language: a systematic review of the machine learning classification," PeerJ Computer Science, vol. 8, p. e883, Feb. 2022. Available: https://doi.org/10.7717/peerj-cs.883.

[7] H. M. Hamza and A. Wali, "Pakistan sign language recognition: leveraging deep learning models with limited dataset," Machine Vision and Applications, vol. 34, no. 5, Jul. 2023. Available: https://doi.org/10.1007/s00138-023-01429-8.

[8] "PSL — Pakistan Sign Language." www.psl.org.pk, psl.org.pk.

Undertaking of Kaavish advisement as an External Supervisor

I hereby affirm that I have read the project details as described on the preceding pages and agree to undertake advisement of this Kaavish project as an External Supervisor. I understand that this role entails the following.

Meeting Meeting the project team regularly, at least once every two weeks, for the entire duration of the Kaavish. The meetings may be held remotely if required.

Advisement Providing supervision and advice to the team in order to ensure steady progress of the project toward its goals.

Liaison Liaising with the Internal Supervisor as required, e.g. to provide feedback or engage in grading.

Other Any other task, depending on availability and suitability, relevant to the Kaavish as communicated by the Internal Supervisor or Kaavish Working Group.

Name: _____

Email: _____

Phone: _____

Designation: _____

Affiliation: _____

Signature: _____