Tamil Alphabet Sign Language Recognition Using Mediapipe and Machine Learning

Vidura Perera
Department of Physical Science
University of Vavuniya
Vavuniya, Sri Lanka
gviduraperera@gmail.com

Vijayakanthan Ganesalingam
Department of Physical Science
University of Vavuniya
Vavuniya, Sri Lanka
g.vijayakanthan@vau.ac.lk

Vaishali Ravi
Department of Physical Science
University of Vavuniya
Vavuniya, Sri Lanka
https://orcid.org/0000-0002-5463-4071

Jeyamugan Thirunavukkarasu
Department of Physical Science
University of Vavuniya
Vavuniya, Sri Lanka
tjeyamugan@vau.ac.lk

2024 9th International Conference on Information Technology Research (ICITR) | 979-8-3315-1142-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICITR64794.2024.10857766

***Abstract*—Tamil is an ancient and unique language with 247 characters, and it can be challenging for the deaf and mute community to communicate its alphabets through sign language to others, often leading to social exclusion. To bridge this communication gap, we propose a real-time Tamil sign language recognition system using the Mediapipe framework and various machine learning algorithms. Our study trained multiple models, including Support Vector Machine (SVM), Random Forest, XG-Boost, and Gradient Boosting, using a publicly available dataset specifically developed for Tamil letters and their corresponding signs. Among these models, the Random Forest classifier demonstrated the best performance, achieving an accuracy of 96.45%. This outcome provides a state-of-the-art, efficient real-time Tamil sign language translator, enhancing communication and promoting inclusivity within the Tamil community.**

Keywords: Machine Learning, Mediapipe, Random Forest, Tamil sign language

I. Introduction

The Tamil language is unique and ancient, dating back to the third and first centuries BC [1]. It is spoken by approximately 85 million people worldwide, making it the 17th most spoken language globally [2], [3]. Tamil is an official language in Sri Lanka, India, and Singapore and is widely spoken in countries such as Malaysia, Fiji, South Africa, and Mauritius [4]. However, within the Tamil-speaking community, the deaf and mute population is often marginalized. A survey conducted in Tamil Nadu revealed that about 3% of the population was identified as mute or deaf [5], and by 2023, 37.46% of the population was recognized as disabled [6]. This increase reveals the growing size of the disabled community in Tamil-speaking regions.

As the population grows, it has become more difficult for deaf and mute people to communicate effectively with the general public since sign language interpretation is not universally understood [7]. This communication barrier contributes to social isolation and shows the importance of a system that can bridge the gap between the deaf and mute and the rest of the community [8]. It is essential to develop a system that can accurately translate Tamil sign language, helping to overcome communication barriers and promote inclusivity.

The Tamil language has 247 characters, consisting of 12 vowels (*uyir ezhuthu*), 18 consonants (*mey ezhuthu*), 216 compound characters (*uyirmey ezhuthu*), and one special character called *ayudha ezhuthu* [6]. The large number of characters makes it difficult for ordinary people to quickly learn and understand Tamil sign language, creating a significant communication barrier for the deaf and mute community [9]. To address this challenge, various techniques using image processing, machine learning, deep learning, and virtual reality have been applied to sign language recognition. Numerous studies have focused on different languages, such as British Sign Language (BSL), American Sign Language (ASL), and other sign languages in the French, Japanese, and Australian language families [10][11]. These studies have successfully demonstrated the potential of technology in improving communication for the hearing and speech-impaired. However, despite the advancements in sign language recognition for other languages, Tamil sign language remains underdeveloped. While several approaches have been introduced, there is still a need for a comprehensive, real-time, and accurate interpreter that can effectively cater to the unique complexity of Tamil characters and facilitate seamless communication.

This study proposes a system that enhances human-computer interaction (HCI) by providing an interface to translate Tamil sign language for deaf or mute individuals. Machine learning techniques are employed in conjunction with the Mediapipe Hand Landmark model (MHL), developed by Google, to detect 21 landmarks on the hand, as shown in Fig. 1 [12]. The MHL model identifies the palm in an image and specific landmarks on the detected palm. The framework uses a bounding box to localize the hand region in one video frame and applies it to subsequent frames to reduce processing time [13]. This optimization reduces the need for hand detection in every frame, improving real-time performance.

Machine learning algorithms, including Random Forest [14], Gradient Boosting [15], XGBoost [16], and Support Vector Machine(SVM) [17], were trained and compared to analyze their performance. The Random Forest model demonstrated superior performance and accuracy compared to the other models. This finding demonstrates the effectiveness of Random Forest in capturing the complex patterns of Tamil sign language. The primary objective of this study is to support deaf and mute individuals in actively interacting within the Tamil-speaking community. The paper is structured as follows: Section 2 discusses existing works and research gaps, Section 3 presents the proposed methodology and technologies, and Section 4 provides results and comparative analysis, and the conclusion outlines future work.

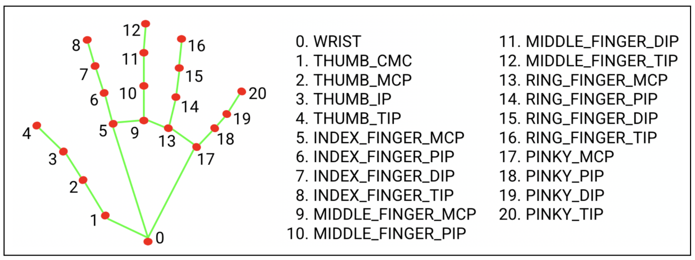

Fig. 1. Key points marked by Mediapipe Hand Landmarks

II. Literature Review

Research on Tamil Sign Language (TSL) is relatively limited compared to existing sign recognition systems. The studies that have been conducted primarily address alphabet recognition, translation to speech, and dynamic recognition systems, each employing distinct methods and achieving varying degrees of success and limitations.

One study [18] introduced a method for converting a set of 32 combinations of binary numbers, representing the ”UP” and ”DOWN” positions of five fingers, into decimal numbers. The Palm Extraction method, feature point extraction, and binary-decimal conversion algorithms were employed to translate these into corresponding Tamil letters, including vowels, consonants, and special characters, achieving an accuracy of 98.75%. However, this approach faced challenges related to background removal in images, which is crucial for accurately identifying hand gestures in different scenes.

Another notable research by Sudha and Jothilakshmi [10] used various classifiers such as K-Nearest Neighbor (KNN), Proximal Support Vector Machine (PSVM), and Naïve Bayesian to translate TSL to speech. The classifiers achieved varying degrees of accuracy, with Naïve Bayesian reaching the highest at 93%. The study aimed to enhance recognition rates by introducing shape descriptors to account for viewpoint variations, but this approach limited the system's scalability and adaptability to different sign languages.

In the domain of dynamic gesture recognition, a study [19] was focused on recognizing dynamic gestures using a vision-based approach. The system employed the Two-Dimensional Discrete Sine Transform (DST) for image compression and a Self-Organizing Map (SOM) neural network for pattern recognition, achieving a 91% accuracy rate. However, this study mainly targeted dynamic gestures and did not adequately address static gestures, which are equally important for effective communication in TSL.

Jayanthi and Thyagharajan [20] applied HCI principles to facilitate communication between the deaf-mute community and non-signers. This research employed the Generalized Hough Transform technique for feature extraction and successfully recognized 31 Tamil language alphabets. Nevertheless, the system can work under a controlled environment with a non-reflective background, which poses a challenge for real-world applications, where such ideal conditions are not always available.

Another approach[21] was introduced by using Convolutional Neural Network (CNN) to assist deaf and mute individuals by converting static and dynamic hand gestures into sound waves. Although Conv2D was used for model training and testing, the study did not detail potential future advancements or the use of advanced technologies for developing faster and more accurate translations.

Furthermore, Priya et al. [6] presented a technique based on edge detection, a vital system component for recognizing TSL. Canny edge detection, two gesture recognition methods, and a scale-invariant feature detection transform algorithm were all used successfully to train the input system. Despite their promise, these approaches could not consistently provide high-quality images, negatively affecting system performance.

A different study investigated the recognition of Sri Lankan TSL, focusing on the finger spelling alphabet [9]. The study employed a neural network trained with more than 300 sample instances of each Tamil finger spelling, achieving an overall recognition accuracy of 73.76%. However, the recognition rate for consonants (74.72%) was higher compared to vowels (71.5%). This result indicates variability in the effectiveness of the recognition system depending on the type of character, suggesting the need for further refinement.

Even though many studies have been carried out related to TSL, there is a significant gap that remains in terms of robustness, which fails to account for both static and dynamic gestures, overly reliant on controlled environments and predefined descriptors. Additionally, many studies haven't used the state-of-the-art tool MHL which helps in enhancing the accuracy and real-time applicability of sign languages. There is a need for a comprehensive solution that integrates modern machine learning techniques and advanced hand-tracking models to deliver a reliable and effective real-time TSL interpreter. Our proposed system addresses this gap by combining machine learning algorithms with the MediaPipe framework, aiming to provide a more inclusive, user-friendly, and accurate tool for the Tamil-speaking deaf and mute community.

III. Proposed Methodology

This section describes the approach used to develop a TSL using MHL and ML techniques. The approach started with data preparation, implementation of hand landmark detection, initial testing and model selection, and the development of a real-time gesture recognition system as shown in Fig. 2.

A. Data Preparation

This study used the TLFS23 dataset [11], which is publicly available on Mendeley Data. The dataset contains 255,155 images across 248 classes. There are 247 classes representing each character of Tamil, along with an additional class that includes images without any hands. Each class contains approximately 1,000 images to train the model for hand detection. This dataset was collected from 120 individuals from 15 to 80 years old ensuring the different physical natures of hand is also can be recognized [11]. For this study, only those 247 classes were used, resulting 254,147 images in total. The MHL framework extracts the hand key points from these images, and the extracted key points are converted into pickle object to train the machine learning models.

B. Mediapipe Hand Landmark Detection

In this research, the MediaPipe framework is utilized for real-time hand gesture recognition due to its efficiency and advanced capabilities. The MediaPipe Hands pipeline comprises two key models that coordinate to facilitate hand gesture recognition. The first is the Palm Detector, which employs a Single Shot Detection (SSD) architecture to identify the presence and location of a hand within the input image. Once a hand is detected, the process advances to the Hand Landmark Model. This model utilizes a regression-based architecture to predict the precise positions of 21 hand landmarks, enabling detailed analysis of hand posture and movement. Together, these models form a robust framework for hand gesture recognition and interaction applications. This framework enables accurate and robust hand tracking using a single RGB camera [8], [12], [13].

This project intends to improve TSL recognition using MediaPipe's pre-trained models, offering a scalable and effective solution for real-time gesture interpretation. By integrating MediaPipe with machine learning models, earlier shortcomings in sign language identification systems are addressed and effective communication between the Tamil-speaking society and deaf and mute individuals is ensured.

After extracting the hand landmarks, the data was saved in the pickle file format [22], a binary protocol unique to Python that is used to serialize and de-serialize Python object structures. In this scenario, the pickle file contains the extracted hand landmarks and labels, which will be used to train a machine learning model.

C. Model Selection

Following the creation of the pickle data file, the next step involved using this data for initial testing and identifying the most suitable classifier. A subset of the dataset containing Tamil vowels was used to evaluate various classifiers. We compared the performance of the classifiers Random Forest, SVM, Gradient Boosting, and XGBoost with default parameter values. By conducting a comparative analysis of these models, the accuracy and efficiency were successfully assessed. The model demonstrating the highest performance was selected for further development and integration into the TSL recognition system.

After the model selection, the performance of the developed system was rigorously evaluated using key metrics: accuracy, precision, recall, and F1-score. These metrics were selected to provide a comprehensive assessment of the system's capability to accurately recognize and interpret TSL gestures across various categories and variations [6], [7], [14], [21]. The evaluation process involved testing the system on a diverse set of gestures, to ensure its robustness and adaptability to the variability encountered in real-world scenarios.

IV. Results and Discussion

A. Experimental Setup

This research is conducted using Python as the primary programming language, incorporating libraries such as *Pickle*, *MediaPipe* [23], *OpenCV* [24], and *Scikit-learn* [25]. Additionally, *Pygame* [26] is integrated to display real-time gesture recognition output by capturing images from input devices and translating and displaying gestures on a user interface. The primary resource for capturing hand gestures is a webcam. Various webcams are used to detect gestures to ensure the model's effectiveness across different image qualities.

B. Model Performance on Initial Testing

To initially select the best classifier for recognizing TSL accurately, the evaluation metrics, including Accuracy, Precision, Recall, and F1-score, were calculated.

1) *Accuracy*: measures the proportion of correctly predicted instances among the total instances evaluated. In our initial tests, the classifier with the highest accuracy was identified as the most suitable for TSL recognition.
2) *Precision*: quantifies the number of correct positive predictions made by the classifier out of all positive predictions. This is crucial for ensuring that recognized gestures are indeed correct, minimizing errors in TSL interpretation.
3) *Recall*: or sensitivity, measures the proportion of actual positive instances that were correctly identified by the classifier. It is critical in TSL recognition as it reflects the model's ability to detect all relevant gestures, especially those that may be less common.
4) *F1-score*: is a balanced statistic that takes into account both false positives and false negatives. In our tests, the F1-score helped determine the classifier that best balances these two aspects, providing a more reliable measure of the model's overall effectiveness.

Table I shows the evaluation metrics for the various classifiers tested on Tamil vowels. Based on the results, the Random Forest model achieved the highest accuracy of 96.81%, outperforming the other models. While the XGBoost model also demonstrated high accuracy at 96.40%, it showed a notable difference in precision, recall, and F1-score, compared to

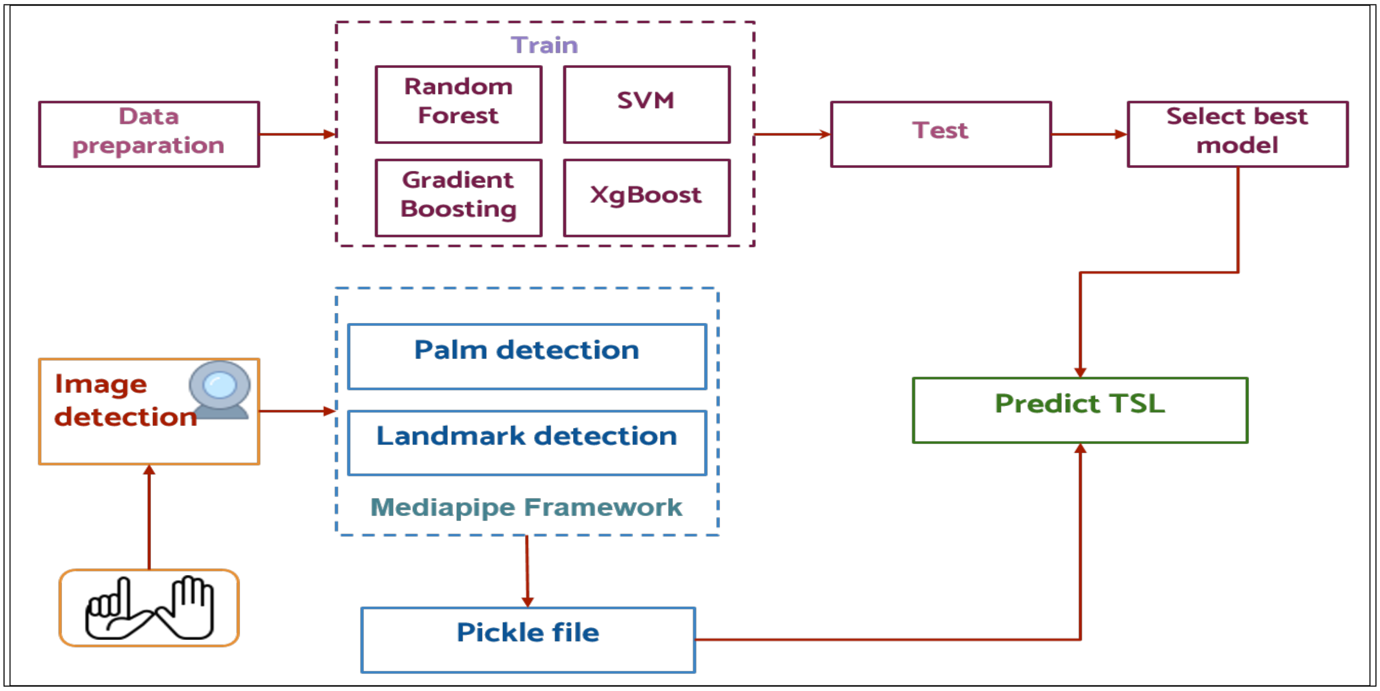

Fig. 2. Workflow of the Proposed Study

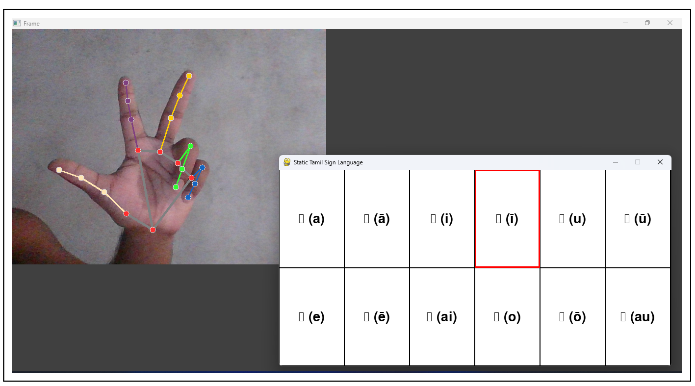

Fig. 3. TSL detection of only Tamil vowels

Random Forest. The Gradient Boosting model achieved an accuracy of 94.63%, and the SVM model had the lowest accuracy at 90.39%. Despite their relatively high accuracy, Gradient Boosting and SVM also displayed lower scores in precision, recall, and F1-score compared to Random Forest.

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT MODELS

Models	Accuracy	Precision	Recall	F1-score
Random Forest	96.81%	0.97	0.97	0.97
XGBoost	96.40%	0.96	0.96	0.96
Gradient Boosting	94.63%	0.95	0.95	0.95
SVM	90.39%	0.94	0.90	0.92

C. *Real-time Prediction*

After selecting Random Forest as the best model, its performance was tested with various cameras and individuals while capturing static gestures. Figure 3 illustrates the static translation of sign language for Tamil vowels, where the model highlights the corresponding Tamil letter on the user interface. In contrast, as shown in Figure 4, the same model was used to predict all Tamil alphabets, achieving an average accuracy of 96.45 %. The system then displays the possible interpretations of the sign language gesture.

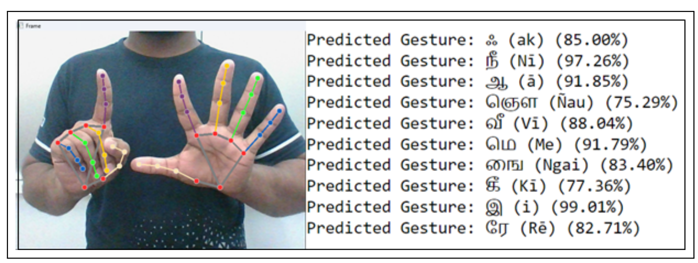

Fig. 4. TSL detection for all letters

V. CONCLUSION AND FUTURE ENHANCEMENT

It is crucial to provide effective support for deaf and mute individuals to facilitate their interaction with the broader community. This study employs machine learning techniques alongside MHL to deliver a real-time, efficient TSL character translator. The system performs well when using the Random Forest algorithm, which yielded an accuracy of 96.45%. Future work will focus on improving accuracy by exploring additional machine learning techniques and expanding the translator's capability to interpret complete Tamil words. Moreover, audio feedback for the translated signs will be integrated in an effort to improve the user experience.

References

[1] V. Naidu, "Tamil, the Indian Language that Spread Across the World," The Juggernaut summit, 2024. [Online]. Available: https://www.thejuggernaut.com/tamil-history-indian-language-movement. [Accessed: Aug. 15, 2024].

[2] B. Krishnamurti, "Tamil Language," Britannica, 2024. [Online]. Available: https://www.britannica.com/topic/Tamil-language. [Accessed: Sep. 05, 2024].

[3] "Identity - Tamil Language," World mapper, 2005. [Online]. Available: https://worldmapper.org/maps/tamil-language-2005/. [Accessed: Aug. 15, 2024].

[4] "List of countries and territories where Tamil is an official language," Wikipedia, 2024. [Online]. Available: https://en.wikipedia.org/wiki/List_of_countries_and_territories_where_Tamil_is_an_official_language. [Accessed: Aug. 15, 2024].

[5] Y. Kabirdoss, "Tamil Nadu govt moves bill to allow deaf and speech impaired people," The Times of India, 2019. [Online]. Available: https://timesofindia.indiatimes.com/city/chennai/tamil-nadu-govt-moves-bill-to-allow-disabled-people-to-contest-local-body-elections/articleshow/70296426.cms. [Accessed: Sep. 02, 2024].

[6] C. B. Priya, S. P. S. Ibrahim, D. Y. Thangam, X. F. Jency, and P. Parthasarathi, "Tamil Sign Language Translation and Recognition System for Deaf-mute People Using Image Processing Techniques," Appl. Comput. Eng., vol. 8, no. 1, pp. 380–386, Aug. 2023, doi: 10.54254/2755-2721/8/20230193.

[7] F. Naseer, G. Ullah, M. A. Siddiqui, M. Jawad Khan, K.-S. Hong, and N. Naseer, "Deep Learning-Based Unmanned Aerial Vehicle Control with Hand Gesture and Computer Vision," in 2022 13th Asian Control Conference (ASCC), May 2022, pp. 1–6. doi: 10.23919/ASCC56756.2022.9828347.

[8] L. Ivanska and T. Korotyeyeva, "Mobile real-time gesture detection application for sign language learning," in 2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 2022, pp. 511–514, doi: 10.1109/CSIT56902.2022.10000440.

[9] S. Vanjikumaran and G. Balachandran, "An automated vision based recognition system for Sri Lankan Tamil sign language finger spelling," in 2011 International Conference on Advances in ICT for Emerging Regions (ICTer), Sep. 2011, pp. 39–44. doi: 10.1109/ICTer.2011.6075024.

[10] S. Sudha and S. JothiLakshmi, "Tamil Sign Language to Speech Translation," Int. J. Comput. Appl., vol. 82, no. 11, pp. 40–45, 2013, doi: 10.5120/14164-2335.

[11] B. R. S, C. M, A. S, G. VA, B. B. Nair, and H. ME, "TLFS23 Tamil language fingerspelling dataset," Data Br., vol. 52, p. 109961, Feb. 2024, doi: 10.1016/j.dib.2023.109961.

[12] Google, "Hand landmarks detection guide." [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker. [Accessed: Jun. 06, 2024].

[13] K. Km and S. Nr, "Recognization of Hand Gestures Using Mediapipe Hands," Int. Res. J. Mod. Eng. Technol. Sci., no. 06, pp. 2582–5208, 2022. [Online]. Available: www.irjmets.com.

[14] N. Bargellesi, M. Carletti, A. Cenedese, G. A. Susto, and M. Terzi, "A Random Forest-based Approach for Hand Gesture Recognition with Wireless Wearable Motion Capture Sensors," IFAC-PapersOnLine, vol. 52, no. 11, pp. 128–133, 2019, doi: 10.1016/j.ifacol.2019.09.129.

[15] J. Nayak, B. Naik, P. B. Dash, A. Souri, and V. Shanmuganathan, "Hyper-parameter tuned light gradient boosting machine using memetic firefly algorithm for hand gesture recognition," Appl. Soft Comput., vol. 107, p. 107478, Aug. 2021, doi: 10.1016/j.asoc.2021.107478.

[16] S. Vostrikov, M. Anderegg, C. Leitner, L. Benini, and A. Cossettini, "Hand Gesture Recognition via Wearable Ultra-Low Power Ultrasound and Gradient-Boosted Tree Classifiers," in 2023 IEEE International Ultrasonics Symposium (IUS), Sep. 2023, pp. 1–4. doi: 10.1109/IUS51837.2023.10307059.

[17] C. Miron, A. Pasarica, H. Costin, V. Manta, R. Timofte, and R. Ciucu, "Hand Gesture Recognition based on SVM Classification," in 2019 E-Health and Bioengineering Conference (EHB), Nov. 2019, pp. 1–6. doi: 10.1109/EHB47216.2019.8969921.

[18] P. S. Rajam and G. Balakrishnan, "Recognition of Tamil Sign Language Alphabet using Image Processing to aid Deaf-Dumb People," Procedia Eng., vol. 30, pp. 861–868, 2012, doi: 10.1016/j.proeng.2012.01.938.

[19] S. Sudha and S. Jothilakshmi, "Dynamic Tamil Sign Language Recognition System," Int. J. Adv. Res. Manag. Archit. Technol. Eng., vol. II, no. VIII, pp. 1–6, 2016.

[20] P. Jayanthi and K. K. Thyagharajan, "Tamil alphabets sign language translator," in 2013 Fifth International Conference on Advanced Computing (ICoAC), Dec. 2013, pp. 383–388. doi: 10.1109/ICoAC.2013.6921981.

[21] S. J. Swetha and S. R. Mary, "Tamil Sign Language Recognition using Machine Learning to Aid Deaf and Dumb People," Int. Res. J. Eng. Technol., p. 3195, 2008. [Online]. Available: www.irjet.net.

[22] Python, "pickle — Python object serialization." [Online]. Available: https://docs.python.org/3/library/pickle.html. [Accessed: Jun. 01, 2024].

[23] Python, "mediapipe 0.10.15." [Online]. Available: https://pypi.org/project/mediapipe/. [Accessed: Jun. 05, 2024].

[24] Python, "opencv-python 4.10.0.84." [Online]. Available: https://pypi.org/project/opencv-python/. [Accessed: Jun. 05, 2024].

[25] Python, "scikit-learn." [Online]. Available: https://scikit-learn.org/stable/. [Accessed: Jun. 05, 2024].

[26] P. Shinners, "Pygame." [Online]. Available: https://www.pygame.org/news. [Accessed: Jun. 05, 2024].