REPUBLIC OF TURKEY

ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

FACULTY OF ENGINEERING

DEPARTMENT OF COMPUTER ENGINEERING

REAL-TIME ASL RECOGNITION USING HAND LANDMARKS

KÜBRASULTAN KÖSE

BACHELOR DEGREE

ADANA 2026

REPUBLIC OF TURKEY
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTER ENGINEERING

REAL-TIME ASL RECOGNITION USING HAND LANDMARKS

KÜBRASULTAN KÖSE
BACHELOR DEGREE

SUPERVISOR
ASSISTANT PROFESSOR DR. BURÇAK ASAL

ADANA 2026

ABSTRACT

REAL-TIME ASL RECOGNITION USING HAND LANDMARKS

Kübrasultan KÖSE

Department of Computer Engineering

Supervisor: Assis. Prof. Dr. Burçak ASAL

01 2026, 29 pages

This study presents a real-time American Sign Language (ASL) alphabet recognition system based on computer vision and machine learning techniques. The proposed system aims to reduce communication barriers between sign language users and non-sign language users by translating ASL hand gestures into textual output in real time.

Live video input is captured using a standard camera, and hand landmarks are detected using the MediaPipe framework. A total of 63 spatial features extracted from hand landmarks are normalized to ensure scale and position invariance. These features are then classified using a Random Forest model trained on a custom-collected ASL dataset. The system is designed to operate efficiently in real time and includes additional mechanisms such as prediction stabilization, confidence-based filtering, and automatic text generation.

Experimental results demonstrate that the proposed approach achieves high classification accuracy on the collected dataset while maintaining stable real-time performance. The findings indicate that landmark-based feature extraction combined with traditional machine learning models can provide an effective and computationally efficient solution for real-time ASL alphabet recognition. This project serves as a functional prototype and a foundation for future extensions toward full sign language translation systems.

Keywords: American Sign Language, Hand Gesture Recognition, Computer Vision, Machine Learning, MediaPipe, Random Forest, Real-Time Systems

ACKNOWLEDGEMENTS

I would like to thank my supervisor proffessor, my friends, and my family for their invaluable feedback and encouragement throughout the development of Real-Time ASL Recognition Using Hand Landmarks.

TABLE OF CONTENTS

LIST OF FIGURES

NOMENCLATURE

ML	: Machine Learning
ASL	: American Sign Language
TSL	: Turkish Sign Language
MCP	: Metacarpophalangeal Joint

1. INTRODUCTION

Sign language recognition systems have gained increasing attention in recent years due to their potential to improve accessibility and inclusivity for hearing-impaired individuals. By combining computer vision techniques with machine learning models, it is possible to automatically interpret hand gestures captured by a camera and convert them into meaningful textual representations. This project focuses on developing a practical and efficient solution that can operate in real-time without the need for specialized hardware.

1.1. Background and Motivation

Sign language is a primary communication method for individuals with hearing and speech impairments. However, the lack of widespread knowledge of sign language among the general population creates communication barriers in daily life. These barriers limit social interaction, access to services, and inclusion in educational and professional environments.

With recent advances in computer vision and machine learning, automatic sign language recognition systems have become an important research area. Such systems aim to translate hand gestures into textual or spoken language, enabling more accessible and inclusive communication between hearing-impaired individuals and the wider society.

Figure 1: ASL Alphabet hand Gestures

1.2. Problem Definition

Despite existing studies on sign language recognition, real-time and user-friendly applications remain challenging due to variations in hand shapes, lighting conditions, camera quality, and individual differences among users. Many systems either require specialized hardware or fail to provide stable real-time performance.

The main problem addressed in this project is the development of a real-time American Sign Language (ASL) alphabet recognition system that operates using a standard camera, provides reliable predictions, and allows users to convert recognized signs into readable text.

1.3. Aim and Scope of the Project

The aim of this project is to design and implement a real-time ASL alphabet recognition system using computer vision and machine learning techniques. The system detects hand landmarks from live camera input, extracts meaningful features, and classifies hand gestures corresponding to ASL letters.

The scope of the project includes single-hand static ASL alphabet gestures, excluding dynamic gestures such as “J” and “Z”. The system focuses on letter-level recognition and provides text output along with optional word suggestions. The project does not aim to perform full sentence-level sign language translation. .

1.4. Contributions of the Project

The main contributions of this project can be summarized as follows:

- Development of a real-time ASL alphabet recognition system using a standard webcam.
- Integration of MediaPipe hand landmark detection for robust and accurate feature extraction.
- Implementation of a machine learning-based classification model for ASL letter recognition.

- Creation of a custom data collection pipeline for building a labeled ASL dataset.
- Design of a user-friendly graphical interface that displays predictions and generated text in real time.

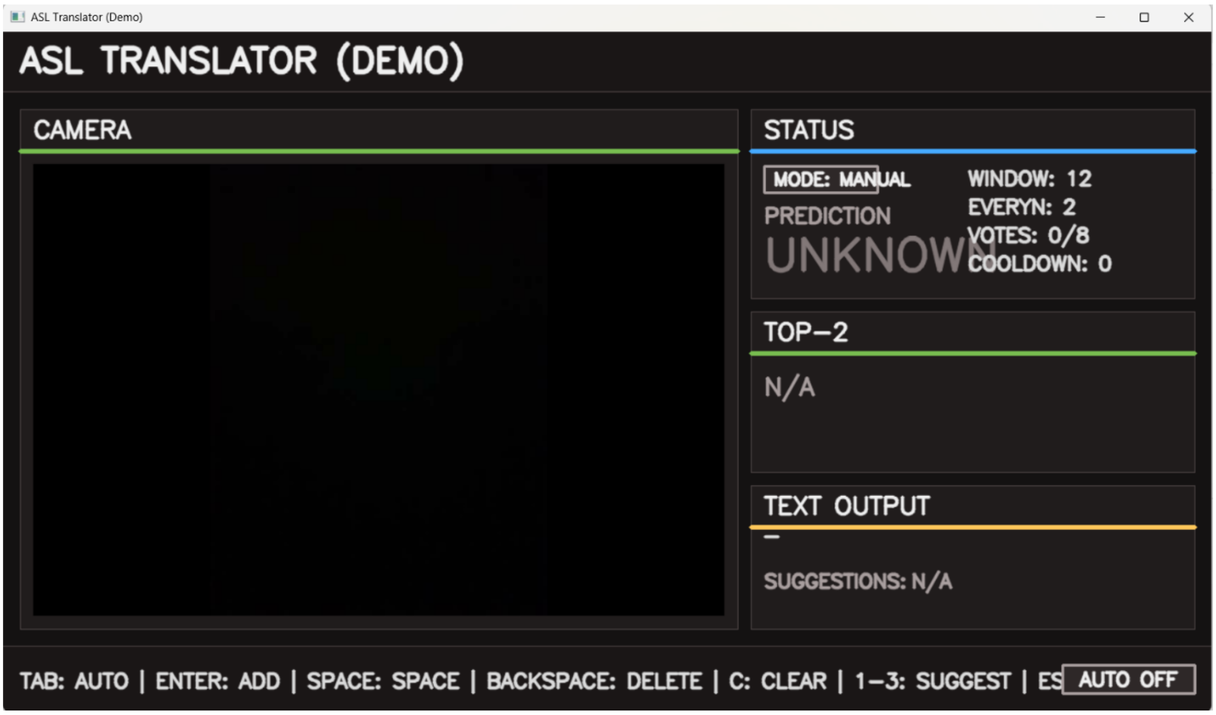

Figure 2: ASL Translator Screen

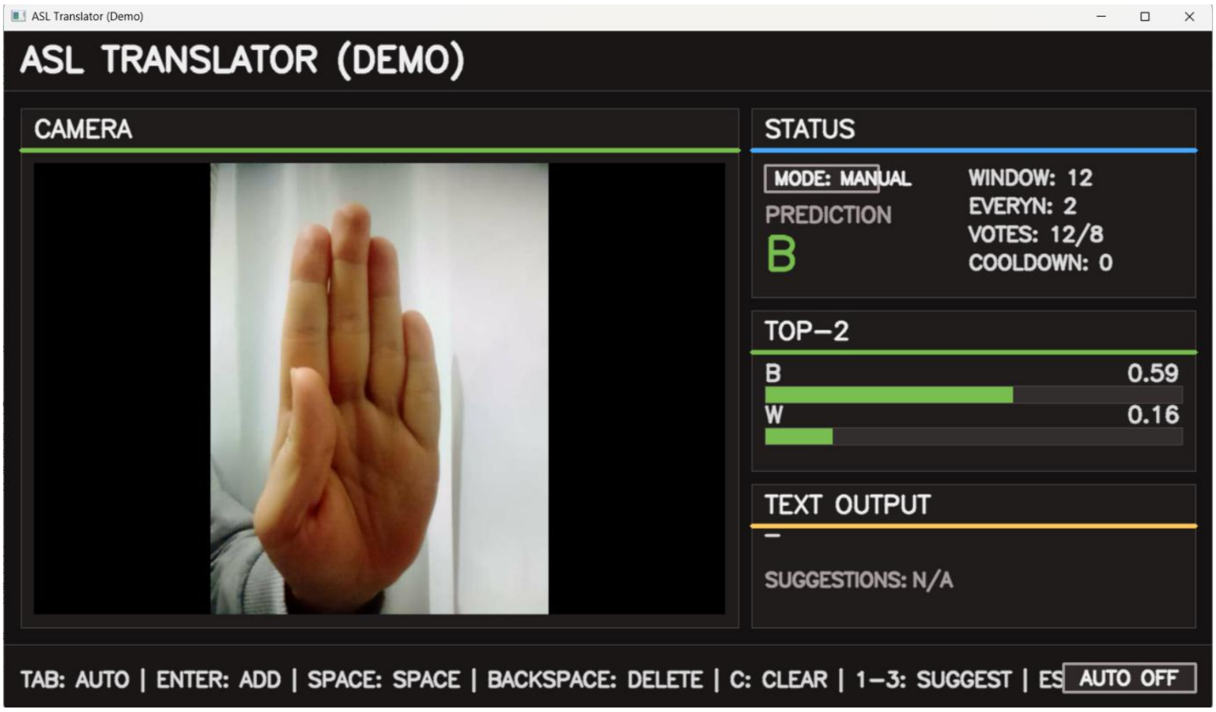

Figure 3: ASL Translator Example

2. RELATED WORK

Sign language recognition has been an active research area for several decades, particularly with the advancement of computer vision and machine learning techniques. Early studies primarily relied on wearable sensors such as data gloves to capture hand movements and finger positions. Although these approaches provided accurate measurements, they required specialized hardware, which limited their usability in real-world applications.

With the widespread availability of cameras and increased computational power, vision-based sign language recognition systems have become more popular. These systems typically utilize image processing techniques to detect hands and extract relevant features such as hand shape, finger orientation, and joint positions. Recent studies have shown that deep learning and classical machine learning models can effectively classify static hand gestures representing sign language alphabets.

In recent years, frameworks such as MediaPipe have significantly improved the robustness of hand detection and tracking. MediaPipe provides real-time hand landmark detection by identifying key joint points on the hand, enabling precise and consistent feature extraction. Several studies have successfully combined MediaPipe hand landmarks with machine learning classifiers such as Support Vector Machines, Random Forests, and Neural Networks to achieve high recognition accuracy for sign language gestures.

Despite the high accuracy reported in controlled environments, many existing systems face challenges in real-time performance, prediction stability, and generalization across different users. Factors such as hand position variation, lighting conditions, and individual differences in gesture execution can negatively affect recognition performance. Therefore, stabilization techniques and user-independent data collection remain important research topics in sign language recognition systems.

In this context, the present project focuses on developing a real-time, camera-based ASL alphabet recognition system that emphasizes usability, prediction stability, and low-cost deployment. By leveraging MediaPipe hand landmarks and a machine learning-based classification approach, the proposed system aims to provide an accessible and extensible foundation for future sign language translation applications.

3. METHODOLOGY

Lorem ipsum dolor sit amet, arcu mauris nostrum magna, justo a unde elementum sapien vitae, et curabitur eget lacus quis magna, massa rhoncus mus justo. Consectetuer id purus pede. Quis per et. Diam suscipit sed sodales non nulla. Suspendisse leo vel, sed dolor at nec, eget sit mi. Arcu ligula tempor venenatis quisque vel nec, aliquet amet erat congue nunc neque amet, ornare faucibus conubia, per magna non mauris malesuada nullam donec.

Morbi tellus odio viverra id, nibh curabitur amet, suspendisse aliquam placerat convallis ridiculus tellus. Mi id nec potenti velit. Vestibulum sapien sed ante nam magna, curabitur aliquam sed mattis, dapibus nulla eleifend praesent turpis justo, nisl nec sed, vel vitae. Ornare sollicitudin libero pellentesque, ut magna, sapien non tincidunt libero, dapibus ultricies est elit orci magna. Sed ultricies risus tincidunt, id libero, luctus pharetra, nulla arcu neque in, arcu ut ligula dui wisi et eget. Sollicitudin eget non non, magna nulla, aliquam semper ac libero vitae.

3.1. System Overview

The proposed system is designed as a real-time American Sign Language (ASL) alphabet recognition application based on computer vision and machine learning techniques. The system captures live video input from a camera, detects the user's hand, extracts discriminative hand landmark features, and classifies the corresponding ASL letter. The recognized letters can then be displayed on the screen and optionally combined into words and sentences.

The overall pipeline of the system consists of five main stages:

1. Video acquisition
2. Hand detection and landmark extraction
3. Feature normalization
4. Machine learning-based classification
5. Prediction stabilization and user interaction

```
Camera Input
     ↓
MediaPipe Hand Landmarks
     ↓
Feature Normalization
     ↓
Random Forest Classifier
     ↓
Prediction Stabilization
     ↓
Text Output
```

Figure 4: Overall System Pipeline

3.2. Video Acquisition

Live video input is captured using a standard webcam or mobile camera connection. Each video frame is processed sequentially to enable real-time performance. To reduce latency and ensure smooth operation, the system processes every N-th frame instead of all frames. This approach balances computational efficiency and recognition accuracy.

Additionally, basic image preprocessing operations such as frame resizing and rotation correction are applied to adapt the input to the camera setup and improve detection stability.

3.3. Hand Detection and Landmark Extraction

Hand detection and tracking are performed using the MediaPipe Hand Landmarker framework. MediaPipe identifies a single hand in each frame and extracts 21 three-dimensional hand landmarks corresponding to key joints and finger positions. Each landmark is represented by normalized x, y, and z coordinates relative to the image frame.

These landmarks provide a compact and expressive representation of the hand pose, allowing the system to capture fine-grained differences between ASL letters without relying on raw image pixels.

Figure 5: Hand Landmarks

Figure 6: Hand Landmarks on example

3.4. Feature Normalization

To ensure robustness against variations in hand position, scale, and distance from the camera, a normalization procedure is applied to the extracted landmarks. First, all landmark coordinates are translated such that the wrist joint becomes the origin. Then, the landmarks are scaled by the distance between the wrist and the middle finger metacarpophalangeal (MCP) joint.

This normalization process makes the feature representation invariant to global hand translation and scale differences, improving the model's generalization across different users.

3.5. Classifiation Model

For classification, a Random Forest classifier is employed. Random Forest is an ensemble learning method that combines multiple decision trees to improve classification accuracy and robustness. It is particularly suitable for this task due to its ability to handle non-linear feature relationships and its resistance to overfitting.

The model is trained on a labeled dataset consisting of normalized hand landmark features collected from multiple ASL users. During training, the dataset is split into training and test sets using a stratified approach to preserve class balance.

3.6. Prediction Stabilization

In real-time recognition scenarios, frame-level predictions may fluctuate due to slight hand movements or temporary detection errors. To address this issue, a prediction stabilization mechanism is implemented using a sliding window and majority voting strategy.

Predictions obtained over a fixed window of recent frames are aggregated, and the most frequent label is selected as the final output if it exceeds a predefined vote threshold. Additionally, confidence-based thresholds are applied to suppress uncertain predictions and output an “Unknown” label when necessary.

3.7. User Interaction and Text Generation

The recognized ASL letters are displayed in a graphical user interface (GUI) in real time. Users can manually confirm predictions or enable automatic input mode to construct words and sentences. The interface also includes optional word suggestion functionality to enhance usability and reduce user effort during text entry.

This design allows the system to function both as a recognition demo and as a practical assistive communication tool.

4. DATASET AND DATA COLLECTION

4.1. Dataset Overview

The dataset used in this study consists of hand landmark features extracted from live video recordings of multiple users performing static ASL alphabet gestures. The dataset covers 24 ASL letters (excluding dynamic letters such as J and Z) and is collected under varying hand shapes and slight pose variations to improve robustness.

All samples are represented as 63-dimensional feature vectors derived from 21 hand landmarks, each described by three-dimensional coordinates. .

4.2. Data Collection Procedure

Data collection is performed using a custom-built recording application. During the collection process, users select a target ASL letter and record multiple frames while holding the corresponding hand pose. Each frame is processed independently, and hand landmark features are extracted using MediaPipe.

To ensure label accuracy, data recording is explicitly controlled by the user. Recording can be started or stopped manually, preventing unintended or noisy samples from being included in the dataset .

4.3. Multi – User Data Collection

To increase the generalization capability of the model, the dataset is collected from multiple participants. Each participant contributes samples for all supported ASL letters. This approach introduces natural variations in hand size, finger proportions, and signing styles.

Collecting data from different users reduces user-specific bias and improves the robustness of the trained model when evaluated on unseen individuals.

4.4. Dataset Balancing Strategy

During data collection, special attention is given to maintaining a balanced dataset. The number of recorded frames per ASL letter is monitored and adjusted to prevent class imbalance. Letters with fewer samples are prioritized during additional recording sessions.

This balancing strategy ensures that the classifier does not become biased toward frequently occurring classes and contributes to more stable prediction performance.

4.5. Manuel Comfirmation Samples

In addition to the main dataset, the system supports optional manual confirmation of recognized letters during real-time usage. When enabled, these manually confirmed samples are stored separately from the primary training dataset.

This design allows future extensions such as incremental learning or personalized model adaptation without affecting the integrity of the original dataset.

5. SYSTEM FEATURES AND UZER INTERACTION

5.1. Manuel and Automatic Input Modes

The system provides two interaction modes: manual and automatic. In manual mode, users explicitly confirm each recognized letter before it is added to the output text. In automatic mode, letters are added automatically when a stable prediction is detected over a predefined duration.

This dual-mode design allows users to choose between higher control and faster interaction depending on their preferences .

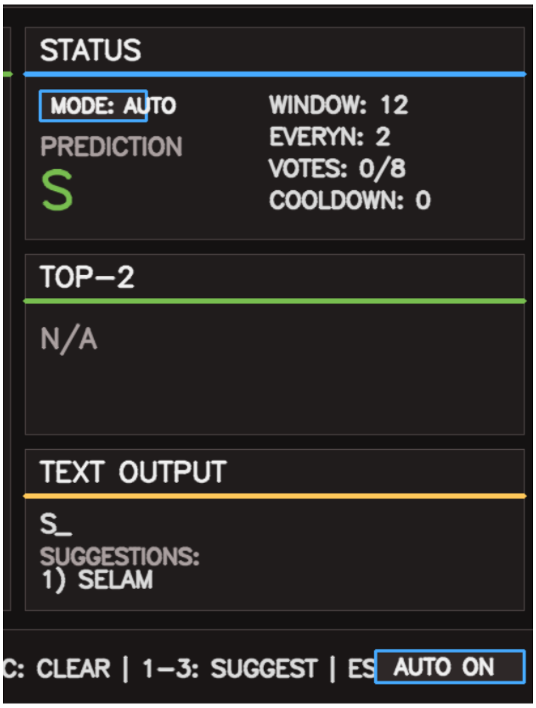

Figure 7: Aotumatic Mode

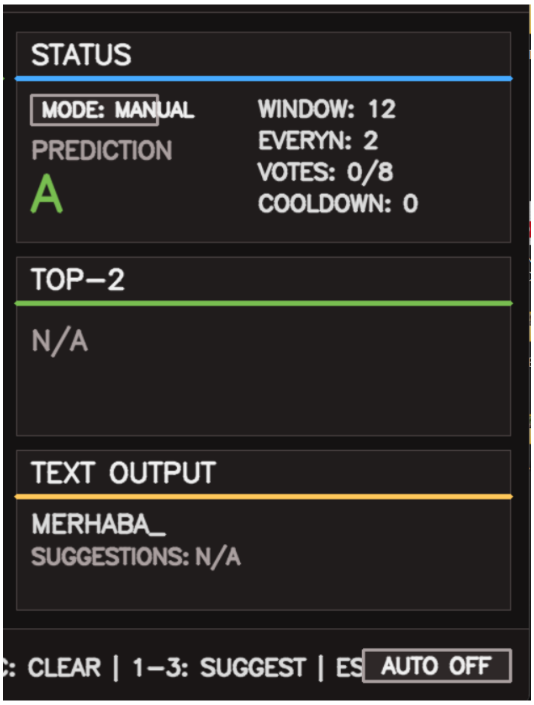

Figure 8: Manuel Mode

5.2. Test Construction and Work Completion

Recognized letters are sequentially combined to form words and sentences. To improve usability, the system includes a word completion mechanism that suggests possible words based on the currently typed prefix.

Word suggestions are displayed in real time and can be selected by the user, reducing the number of required gestures and improving communication efficiency.

5.3. Handling Uncertain Predictions

When the system detects insufficient confidence or unstable predictions, an “Unknown” label is produced instead of forcing an incorrect classification. This mechanism prevents erroneous outputs and improves overall reliability, especially during hand transitions or occlusions.

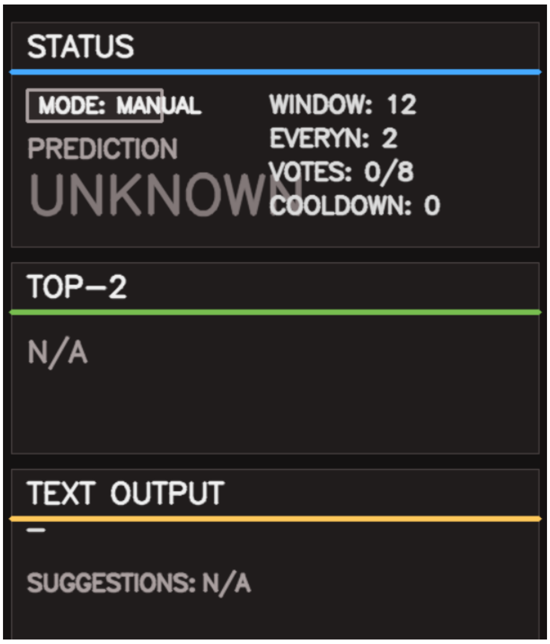

Figure 9: Unknown Label

6. EXPERİMENTAL RESULTS

Accuracy: 0.9990

Classification report:

	precision	recall	f1-score	support
A	1.00	1.00	1.00	256
B	1.00	1.00	1.00	233
C	1.00	1.00	1.00	250
D	1.00	1.00	1.00	240
E	1.00	1.00	1.00	248
F	1.00	1.00	1.00	262
G	1.00	1.00	1.00	257
H	1.00	1.00	1.00	231
I	1.00	1.00	1.00	228
K	1.00	1.00	1.00	245
L	1.00	1.00	1.00	223
M	1.00	1.00	1.00	251
N	1.00	1.00	1.00	265
O	1.00	1.00	1.00	260
P	0.99	1.00	0.99	253
Q	1.00	1.00	1.00	231
R	1.00	0.99	0.99	259
S	1.00	1.00	1.00	257
T	1.00	1.00	1.00	83
U	0.99	1.00	1.00	277
V	1.00	1.00	1.00	293
W	1.00	1.00	1.00	231
X	1.00	1.00	1.00	246
Y	1.00	1.00	1.00	264
accuracy			1.00	5843
macro avg	1.00	1.00	1.00	5843
weighted avg	1.00	1.00	1.00	5843

Figure 10: Classification Results

6.1. Experimental Setup

The trained ASL recognition model is evaluated using a hold-out test set that is separated from the training data. The dataset is randomly split into training and testing subsets with a ratio of 80% for training and 20% for testing while preserving class distribution.

The evaluation is performed on static ASL alphabet gestures represented by normalized 63-dimensional hand landmark feature vectors. All experiments are conducted on a personal computer using real-time webcam input and offline test samples.

6.2. Classification Accuracy

The trained Random Forest classifier achieves an overall classification accuracy of 99.9% on the test dataset. This high accuracy indicates that the model successfully learns discriminative patterns from normalized hand landmark features for static ASL letters. Although the model achieves very high offline accuracy, real-time prediction performance is affected by hand pose transitions and visual similarity between certain ASL signs.

The strong performance is mainly attributed to:

- Consistent feature normalization,
- Balanced class distribution,
- Clear separation between static ASL hand poses.

6.3. Precision, Recall and F1-score Analysis

Detailed performance metrics including precision, recall, and F1-score are computed for each ASL letter. The classification report shows that nearly all classes achieve precision and recall values close to 1.00.

This result demonstrates that the model produces very few false positives and false negatives across different ASL letters, indicating stable and consistent classification behavior.

6.4. Confusion Matrix Evaluation

The confusion matrix reveals that misclassifications between ASL letters are extremely rare. Most predictions are located along the main diagonal, confirming that the classifier correctly distinguishes between visually similar hand shapes.

Minor confusions occur only between a small number of letters with similar finger configurations, which is expected in static hand gesture recognition tasks.

```
Confusion matrix:

[[256   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0]
 [  0 233   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0]
 [  0   0 250   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0]
 [  0   0   0 240   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0]
 [  0   0   0   0 248   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0]
 [  0   0   0   0   0 262   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0]
 [  0   0   0   0   0   0 257   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0]
 [  0   0   0   0   0   0   0 231   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0 228   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0 245   0   0   0   0   0   0   0   0
    0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0 223   0   0   0   0   0   0   0
    0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0 250   0   0   1   0   0   0
    0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0 265   0   0   0   0   0
    0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0 260   0   0   0   0
    0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0 253   0   0   0
    0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 231   0   0
    0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   2   0 256   0
    0   1   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 257
    0   0   0   0   0   0]
```

Figure 11: Confision Matrix

6.5. Real – Time Performance Observation

In addition to offline evaluation, the trained model is tested in a real-time ASL-to-text application. While offline accuracy remains very high, real-time performance is affected by factors such as hand motion, lighting conditions, and transition frames between gestures.

To mitigate these issues, temporal stabilization techniques and confidence-based filtering mechanisms are applied, resulting in improved robustness during live usage.

6.6. Discussion of High Accuracy

Although the reported test accuracy is very high, this result does not imply perfect real-world performance. The controlled nature of the dataset and the use of static gestures simplify the classification task compared to continuous sign language recognition.
Therefore, real-time prediction results are evaluated qualitatively alongside quantitative metrics to provide a more realistic assessment of system performance.

7. DISCUSSION

7.1. Interpretation of Results

The experimental results demonstrate that hand landmark–based features combined with a Random Forest classifier are highly effective for recognizing static ASL alphabet gestures. The achieved classification accuracy and class-wise performance metrics indicate that the model can reliably distinguish between different hand shapes under controlled conditions.

However, despite the high quantitative performance, real-time predictions are more sensitive to variations such as hand movement, slight pose changes, and transition frames between consecutive gestures. These factors introduce ambiguity that is not fully represented in the offline test dataset.

7.2. Limitations of the Current System

The proposed system is designed for recognizing static ASL letters and does not support dynamic gestures such as letters that require motion (e.g., J and Z). Additionally, the system assumes the presence of a single hand and a relatively stable background.

Another limitation is that the model is trained primarily on data collected from a limited number of users. As a result, variations in hand size, finger flexibility, and signing style may negatively affect generalization when the system is used by unseen individuals.

7.3. Real – Time Challenges and Practical Observations

During real-time experiments, misclassifications are observed mainly during gesture transitions or when the hand is partially visible. To address this issue, confidence thresholds, majority voting over multiple frames, and temporal stabilization mechanisms are applied.

These strategies reduce rapid prediction fluctuations and improve user experience, although they may introduce a slight delay in character recognition. This trade-off between responsiveness and stability is a common challenge in real-time vision-based systems.

7.4. Summary of Discussion

Overall, the proposed ASL recognition system demonstrates strong performance for static hand gesture recognition and provides a solid foundation for more advanced sign language translation systems. While certain limitations remain, the current implementation successfully balances accuracy, robustness, and real-time usability.

8. FUTURE WORK

Future improvements of this project can focus on extending both the recognition capability and practical usability of the system.

- Future versions may include support for dynamic ASL letters and gestures that require motion, such as J and Z, by incorporating temporal modeling techniques.
- Extending the system to detect and process both hands simultaneously would enable recognition of more complex signs and improve coverage of full ASL vocabulary.
- Collecting data from a larger and more diverse group of users can improve the generalization capability of the model and reduce performance variations across different hand shapes and signing styles.

- More sophisticated machine learning or deep learning models, such as temporal neural networks, could be explored to capture sequential patterns and improve robustness under real-world conditions.
- Beyond individual character recognition, future work may focus on word- and sentence-level sign language translation by combining gesture recognition with natural language processing techniques.
- The system can be adapted into a web-based or mobile application to improve accessibility and enable broader real-world usage.
- Adapting the system for different sign languages (e.g., TSL)
- Adding language models for smarter word and sentence completion.

9. CONCLUSION

In this study, a real-time American Sign Language (ASL) character recognition system was developed using hand landmark detection and machine learning techniques. The proposed system successfully recognizes static ASL letters from live camera input and converts them into textual output through a user-friendly interface.

Hand landmarks were extracted using MediaPipe, and normalized geometric features were used to train a Random Forest classifier. The model achieved high classification accuracy on the collected dataset, demonstrating that landmark-based representations are effective for static sign language recognition. To improve real-time stability, majority voting, confidence thresholds, and margin-based decision mechanisms were applied.

In addition to character recognition, the system includes practical features such as manual and automatic input modes, text visualization, and word suggestion functionality. These components enhance usability and demonstrate how sign language recognition can be integrated into an interactive application rather than remaining a standalone classifier.

Overall, the results show that the proposed approach provides a reliable and extensible foundation for real-time ASL recognition systems and can serve as a basis for more advanced sign language translation applications.

9.1 Closing Paragraph

This project demonstrates that accurate and responsive sign language recognition can be achieved using lightweight computer vision and machine learning techniques. By focusing on static gestures and real-time performance, the system balances accuracy, efficiency, and usability. The developed framework is modular and open to future extensions, making it suitable for further research and practical applications in assistive communication technologies.

REFERENCES

[1] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf

[2] Lugaresi, C., Tang, J., Nash, H., et al. (2019). MediaPipe: A framework for building perception pipelines.

https://arxiv.org/pdf/1906.08172

[3] iStock. (2018). American Sign Language hand gestures.

Retrieved from
https://www.istockphoto.com/tr/fotoğraf/amerikan-i%C5%9Faret-dili-gm960235324-262223334

[4] Sign Language Recognition Using Hand Landmarks
Kumar,P.,&Singh,R.(2019).
Hand gesture recognition for sign language using machine learning. International Journal of Computer Applications, 178(7), 1–5.