MATRIX Academic International Online Journal of Engineering and Technology
Volume VIII, Issue I, ISSN: 2348-3326

A Helping Hand: Gestures Recognition using Android Mediapipe

Shivam Roy[1] and Anuj Singh[2]
[1,2]Department of Computer Science & Engineering, Sharda School of Engineering and Technology, Sharda University, Greater Noida, India

[1]2022306140.shivam@ug.sharda.ac.in,[2]2022405043.anuj@ug.sharda.ac.in

Abstract. This paper presents a work related to multimodal automatic gestures recognition during human interaction. There was a dedicated database prepared; participants completed tasks based on a command-based structure to realize eight different emotional states. There were three primary feature extraction methods adopted in the system: facial expression recognition, gesture analysis through MediaPipe, and acoustic analysis. MediaPipe, which runs on machine learning for gesture tracking and detection, was crucial for hand movement analysis. It used algorithms like CNNs for key hand landmarks' detection in making the gesture recognition process more accurate. Then, it applied a Bayesian classifier for automatically classifying the emotions based on data. Three types of data were tested: unimodal (single input), bimodal (two inputs), and multimodal (all three inputs together). This was either pre or post-classification. The outcome of these experiments was that multimodal fusion improved recognition rates by more than 10% compared to the best unimodal system. Of these combinations, 'gesture-acoustic' proved most effective. The use of all three types of data resulted in further improvements above the best bimodal combination.

Keywords: Gesture Recognition, Facial Expression Recognition, Bayesian Classifier

1. Introduction

Human-to-human communication comes through with multiple modes of emotions. At times, a sign or message becomes ambiguous in the absence of one modality. For example, sometimes an intended emotion can't be clearly ascertained due to the absence of some crucial visual cue while signing. This happens when the person experiencing the emotion thinks others know all the modalities, say facial expressions or gestures, but perhaps the other person is unaware of all of them. In this context, I am implementing the recognition system using MediaPipe in order to capture and analyze these appropriately.

A gesture combined with a neutral or incompatible facial expression may sometimes result in ambiguity, especially if emotional signals of the facial expression are ambiguous or indiscernible. If the interaction is established based on human sign language recognition, then correct emotion identification will not be guaranteed. For humans, a good system should have emotional intelligence; that is, it should be able to perceive, understand, and respond appropriately to emotions. The interaction is rather more natural if the machine could go into such emotionally perceptive interactions. From

 DOI No: 10.21276/MATRIX.2025.8.1.1

that viewpoint, it becomes more enjoyable to use. The machine needs to recognize the emotional state of its user in order to ascertain the user's communication. It can develop a better understanding of the meaning of what is conveyed by the user via capability for emotional expressiveness and intent. For example, it might even aid the system in adjusting the response based on the satisfaction or dissatisfaction reflected by the user. In this application, I utilize MediaPipe to seize and analyze gestures and facial expressions, thereby creating an interaction that is much more sensitive to such emotional cues.

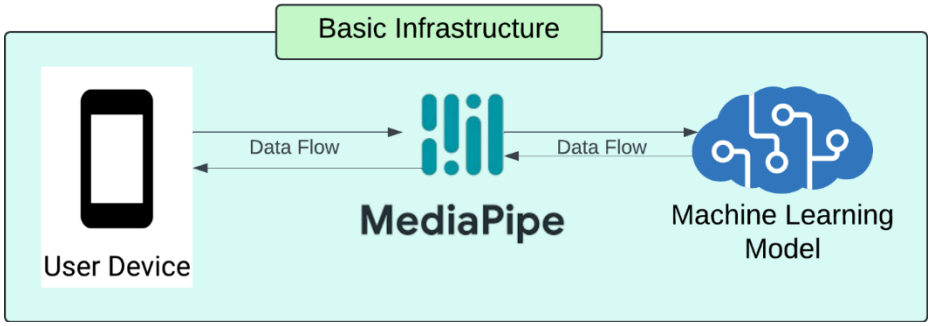

Fig. 1. Basic Infrastructure of Project

Most emotion-analyzing systems focus on just one kind of information at a time, like facial expressions or gestures, but not together. More precisely, there has been very little effort toward bringing together information from body movements and gestures, particularly with regards to sign language. However, a few researchers, such as Sebi et al. and Pantic et al., have reported that the best approach to automatically detecting emotions is through multi-modal input systems just like compound impressions our senses make and, finally, psychological studies show how the merge of behavior from different types can enhance communication by better understanding.

This paper establishes a multimodal approach to recognizing eight different emotional states that may arise during communication, namely Anger, Despair, Interest, Pleasure, Sadness, Irritation, Joy, and Pride. It uses facial expressions with gesture-based information. A Bayesian classifier was trained and then tested with a specially collected dataset made for this study.

The significance of this research is that it combines two different kinds of modalities, namely gestures and facial expressions, in recognition tasks. Because it is a bimodal approach, it derives its information from both kinds of data to achieve a higher accuracy in the interpretation of human emotional expressions of these two modalities, is also examined. While there have been efforts to build systems using two modalities (e.g., combining facial expressions with acoustic data or body gestures [7, 8]), this work specifically targets the underexplored area of integrating sign language gestures with facial expressions.

A further contribution is the incorporation of sign language gestures into a framework for emotion recognition, which has not been studied as extensively as facial expressions. Though gesture has been used to infer emotions in several studies [10–12], integrating gesture with other modalities, such as facial expression, remains less explored.

This work also utilizes features that capture how emotional expressions vary over time. Results suggest that traditional statistical features may not be as effective at

discriminating emotions compared to those that account for the timing and dynamics of facial expressions and gestures.

This paper aims at categorizing various expressions through gestures and face-cue expressions. The performances between the unimodal, bimodal, and multimodal systems are compared as well. I would expect that integration of two types of data streams can yield a higher recognition rate than using a single type of data stream. Results indicate a significant improvement in the performance of the classifier by integrating gestures and face-cue expressions. Specifically, the bimodal fusion of those modalities enhances accuracy more than unimodal systems. The rest of this paper overview the latest development in gesture recognition, explains the data collection process as well as features extracted, and then we will present our proposed approach, starting with a description of each type of data in isolation and then how they are going to be combined towards recognition.

2. Literature Review

Gesture recognition has been heavily pursued in the affective computing domain, with most systems focusing on combining modalities such as facial expressions, speech, and body gestures. It is noteworthy to see that despite the importance of combining sign language with facial expression for emotion recognition, it still remains underexplored in non-verbal communication systems. Facial Expression Recognition is one of the most researched modalities in the detection of emotions. One of the first works done by Ekman and Friesen [1] provided a foundation for most of the modern recognition systems of facial expressions. The later works concentrated on the application of facial expressions to automatic classification of emotions; the development in this area has been very intense both in feature extraction techniques like Active Appearance Models and convolutional neural networks as well as classification Techniques-Support Vector Machines-Bayesian classifiers. However, most of such work has focused on spoken interactions and is still to be replicated in real-world conditions using sign language.

Gesture-Based Emotion Recognition has also made tremendous progress since several works have shown that body movement and hand gestures become a significant part of the expression of emotions. For example, Karg et al. [2] has managed to contribute some portion to gesture-based emotion recognition by using motion capture data to prove that different emotional states are realized through certain gesture patterns. However, Pantic et al. [3] work was quite influential in that it highlighted the necessity of combining facial expression with body gestures in more advanced emotion recognition systems. However, these systems make use of general gestures and not sign language; hence, their applicability is limited to sign language users.

Next to none has addressed the issue of emotion recognition in sign language. In sign language communication, face has an important role, not just in the expression of emotions but also in modulating linguistic elements such as grammar and emphasis. Recent work by Benitez-Quiroz et al. [4] discuss American Sign Language (ASL), identifying patterns that are both linguistic and emotional in content. Despite the significant strides made by this research, the association of the sign language gesture with facial expression for emotion recognition is still limited. Most studies do not join these modalities in a holistic manner. Multimodal Emotion Recognition Systems

have proven that several modalities could be brought together to enhance the detection of emotion. Sebe et al. and Pantic et al. have claimed that an ideal system for affect recognition should be multimodal, where facial, vocal, and gestural information is fused to mimic the human sensory system. These studies have demonstrated that fusion of modalities like facial expressions and speech [7] significantly improves the accuracy over a unimodal system. To this date, the integration of sign language gestures and facial expressions for emotion recognition has received sparse attention.

One of the very few works that specifically deal with multimodal approaches, incorporating gestures as well as facial expressions, is that of Karpouzis et al. [8], which proposed a system for integrating facial, vocal, and bodily expressions for affective state modeling. The work that appears here lends utility to underpin the necessity of multimodal approaches but does not uniquely focus on the special dynamics involved in sign language communication where gestures and facial expressions are intimately interlocked. In HMI systems, emotion recognition plays an increasingly important role for developing more intuitive and adaptive systems. Researchers such as Zeng et al. [9] have proposed multimodal systems that use various combinations of modalities to determine emotions in real-time interactions. Such systems are highly relevant to affective computing but have not been adapted regularly for use in the environment of sign language communication. Direct work on multimodal emotion recognition, especially in sign language contexts, is very sparse. This study aims at filling such a gap by formulating a framework that combines facial expressions and sign language gestures toward improving the accuracy and robustness of emotion recognition systems in non-verbal communication scenarios.

2.1 Collection of Multimodal Data

The collection of multimodal data is a crucial step for building robust emotion recognition systems, particularly for applications involving sign language and facial expression recognition. The data needs to capture the various modalities—gestural, facial, and potentially acoustic cues—in a synchronized manner to effectively model emotional expressions during communication. For this research, we focus on capturing sign language gestures and facial expressions in a structured and controlled environment.

2.2 Participants

The dataset is constructed keeping in mind the fact that participants come from a diverse group of people who are highly trained in the use of sign language, providing representation across gender, age groups, and native signers of different sign languages like American Sign Language (ASL), British Sign Language (BSL), etc. The participants are required to express a range of pre-defined emotions, namely Anger, Sadness, Joy, and Disgust using signed gestures with facial expressions. Since the same set of emotions can be utilized along with both face and sign language, the participants are allowed to express it freely.

2.3 Emotion Categories

The emotional states considered in the dataset are drawn from common affective categories based on Ekman's universal emotions framework and extended to suit non-verbal communication contexts. The emotions include:

Pride, Anger, Sadness, Joy, Surprise, Fear, Disgust, Interest

hese emotions are selected because they cover a wide spectrum of both positive and negative emotional states, making the system more adaptable to various human interaction scenarios.

2.4 Modalities Captured

For a high-quality dataset, this includes different kinds of data collection:

- *Gestures:* The data include hand movements, their direction, and the position of the gestures. Body posture is captured through motion capture technology, including Microsoft Kinect or wearable sensors. In the absence of such, high-definition video cameras are used instead. The recording must be in 3D space to capture all the subtleties of sign language.
- *Facial Expressions:* High-definition cameras record facial movements to capture emotions transferred through the face. Facial Action Units (FACS) are extracted using computer vision techniques, which identify changes in facial muscles and expressions (e.g., eyebrow raises, lip corners, frowns). This information is critical for understanding how emotions are conveyed alongside sign language.
- *Acoustic Signals (Optional):* For emphasis, while the major speech will be in sign language and facial expression, optional acoustic data may be gathered for purposes of including supplementary emotional cues wherever available, such as for non-verbal noises such as sighs, laughter, or gasps.

2.5 Data Synchronization and Annotation

All modalities (gestural, facial, and optional acoustic) are recorded simultaneously to ensure proper synchronization. This is essential for accurate emotion recognition, as timing plays a key role in the perception of emotional cues. Post-processing ensures that the captured data across all modalities is synchronized and aligned temporally.

The captured data is then annotated by human experts about the emotions portrayed. This annotation ranges from marking the emotional content being displayed through gestures and facial expressions with predefined emotion labels. Annotations may further include the intensity of the emotion as well as co-occurring expressions.

2.6 Data Diversity and Variability

To make the model robust to different real-world conditions, the data collection includes variability in terms of lighting, background, and participant diversity. Some sessions are conducted in natural environments, while others take place in controlled lab settings. This diversity helps ensure that the system can generalize well to different users and environments, making it suitable for broader applications in Human-Computer Interaction (HCI) or assistive technologies.

2.7 Multimodal Corpus Creation

The final dataset combines the captured data into a multimodal corpus that can be used for training machine learning models. This corpus includes synchronized recordings

of gestures, facial expressions, and any auxiliary acoustic signals, along with detailed annotations. The dataset is split into training, validation, and testing sets to evaluate the performance of the emotion recognition system.

Table 1. Acted Emotions and Corresponding Emotion-Specific Gestures in Facial Expressions

Emotion	Gesture Description	Facial Expression Characteristics
Anger	Sharp, forceful hand movements; exaggerated signs with strong arm movements.	Furrowed brows, clenched jaw, tense lips, narrowed eyes.
Sadness	Slow, downward hand gestures; signs performed with a slumped posture or lowered shoulders.	Drooping mouth corners, lowered gaze, slightly teary eyes, frown lines.
Joy	Upward, lively hand gestures; fluid, expansive movements with open palms.	Wide smile, raised eyebrows, bright eyes, relaxed facial muscles.
Surprise	Quick, sudden hand movements; abrupt stopping of sign followed by a wide gesture.	Raised eyebrows, widened eyes, open mouth.
Fear	Hesitant, defensive hand movements; gestures closer to the body, as if shielding oneself.	Widened eyes, raised eyebrows, slightly open mouth, tense facial muscles.
Disgust	Quick, dismissive hand flicks; pulling gestures away from the body, as if rejecting something.	Wrinkled nose, furrowed brows, curled upper lip, squinting eyes.
Interest	Smooth, engaged hand movements; signs performed with forward body posture indicating focus.	Raised eyebrows, slight smile, attentive gaze, relaxed forehead.
Pride	Broad, deliberate hand movements; chest-out posture with upward hand motions indicating confidence.	Slight smile, chin raised, eyes looking forward confidently.

3. Feature Extraction

3.1 Face Feature Extraction

The facial feature capturing procedure starts with the face detection and finding its position and boundaries. We used the Viola-Jones algorithm, founded on a sequence of steps to recognize features using patterns called Haar-like features, so helps to approximate the rotation of the head and finds a line between the eyes; therefore, we can rectify the face so this line will be vertical, then the face will be easier for further analysis.

With the face detected, the face is segmented into regions that carry the facial features like eyes, eyebrows, nose, and mouth. This narrows it down and hastens the process. For every feature applied in facial recognition, there has to be a different mask or outline to work perfectly under different lighting conditions. Such masks are then merged into one fine-tuned with human face measurements for accuracy.

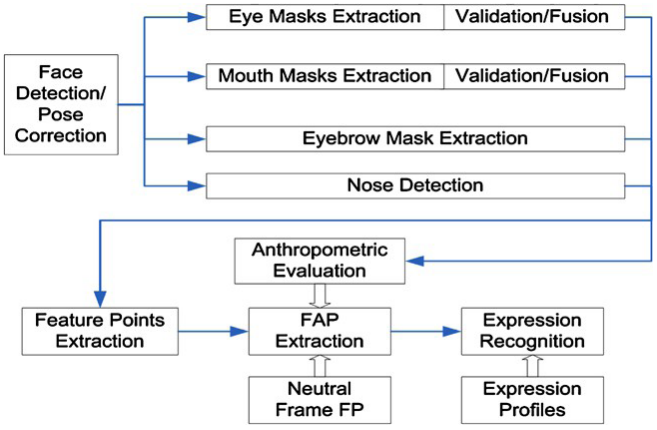

Fig. 2. Face Feature Extraction

We used MPEG-4 Facial Animation Parameters (FAPs) instead of Action Units (AUs) to measure facial deformations. The FAPs were derived from a neutral frame picked out of video sequences. 19 Feature Points (FPs), which could be extracted from the facial region and compared with those of the neutral frame to estimate deformations, produce FAPs used for facial expression estimation.

We process facial features to produce one vector of values per sentence, calculating statistical features over time from these FAP values. Very common challenges include connections with other feature boundaries and mask dislocations due to noise. Mask fusion addresses these issues and enhances the robustness of the approach against variations of lighting conditions and backgrounds.

To validate the facial feature extraction algorithm, we manually annotated 250 randomly chosen frames. To assess agreement between computer generated masks and human observers, we calculated Williams's Index (WI). A WI of more than 1 would indicate better agreement. We obtain values of about 0.838 for the left eye, 0.875 for the right eye, 0.780 for the mouth, and about 1.0 for both eyebrows.

4. Gestures Feature Extraction

Hand position and movement feature extraction in video frames begins with the detection of the hand's position and movements. We depend on a simple, powerful hand detection algorithm, which correctly detects the location and boundaries of hands in the foreground, crucial for capturing all movements related to sign language. The hand gesture interpretation system inducts the detection system that could potentially differentiate hand orientation and position for the estimation of hand gestures which would specifically correspond to signs while tracking their precise gesturally. Once a hand is detected, we then focus on the segmentation of the hand region and extract relevant features to the sign language. We do this by focusing on the important areas within the hand-a shape, palm orientation, and placement of finger. Utilizing anthropometric measurements, we narrow the candidate feature areas that reduce the search space thus hastening the extraction process. For feature extraction, we use the multi-cue approach whereby we apply various algorithms, which strangely happen to have the best performances under different lighting conditions and backgrounds to produce multiple

masks per hand feature. We consider Dynamic Hand Gestures and Static Hand Shapes for sign language. Dynamic gestures capture information over time, considering the trajectory and speed of hand movements. On the other hand, static shapes analyze the configuration of the fingers and the hand position. These features are extracted based on the calculation of several statistical measures over time that capture the evolution of gestures in terms of speed, direction, and spatial orientation.

Another method of feature extraction falls into this category, which is facial expressions. Facial expressions are also important in indicating context and emotion in sign language. Techniques used for tracking facial expressions are similar to those used to track face features extraction, where it tracks key facial points and computes FAPs. We merge the results from the different detection algorithms using mask fusion, to make the final system robust against common challenges encountered, such as occlusion and changing conditions of illumination. This ensures that accuracy is maintained over a dynamic background and against challenging lighting conditions.

In addition, the performance of the proposed sign language feature extraction algorithm is further validated with respect to its effectiveness by introducing manual annotations of selected frames compared to human observer assessments. A measure in terms of the Williams's Index (WI) is also calculated to quantify the agreement between the algorithm's output and human annotations to ensure that the system can correctly recognize and interpret sign language gestures in real-world scenarios.

5. A Framework for Emotion Recognition using Multiple Modalities

In order to test the performance of unimodal, bimodal, and multimodal systems, we employ a standard method by using a Bayesian classifier in the form of BayesNet from the Weka software package. Weka is an open-source toolkit that contains several algorithms for machine learning that can be applied to data mining. The first algorithm we make use of is a Bayesian network, with the 'SimpleEstimator' to derive the conditional probability tables of the network once its structure is set. SimpleEstimator estimates the probabilities directly from the data, with the Alpha parameter set to 0.5 and acting as the initial count for each value in the probability tables. For learning the network structure, we used the K2 learning algorithm a hill-climbing approach that is restricted by an order on the variables as introduced by Cooper and Herskovits. It is a Naïve Bayes Network, therefore it directly connects all other nodes to the classifier node.

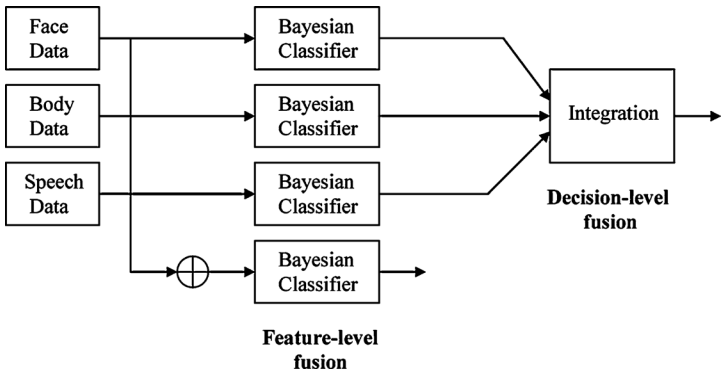

Fig. 3. Overview of the Framework

Figure 3 represents the framework, where we have on the left-hand side, three different Bayesian classifiers for the three modalities: facial expressions, gestures and speech. We also normalized all the datasets; in Weka, there is a normalization function that we used. To ease learning, we applied feature discretization based on Kononenko's Minimum Description Length criterion. This reduces the complexity of the learning task.

We employed a wrapper approach called "WrapperSubsetEval," in which we evaluate different subsets of attributes using a learning scheme to determine which features improve the performance of the classifier. In order to estimate the accuracy of the learning scheme with the selected features, we used cross-validation, with a 5-fold setting, meaning that the data is split into five parts so that we can determine how well the model performs on different subsets. We have set a seed of 1 to make the random splits reproducible. At the end, if the standard deviation of the mean accuracy is more than 0.01, we will reconsider and come back with this judgment again to guide the selection, we developed a best-first search approach in the forward direction. Besides, we trained and tested all of the systems with a 10-fold cross-validation method for robust performance evaluation.

Table 2. Confusion Matrix of the Emotion Recognition System Based on Facial Expressions

a	b	c	d	e	f	g	h	Emotion
56.67	3.33	3.33	10	6.67	10	6.67	3.33	a *Anger*
10	**40**	13.33	10	0	13.33	3.33	10	b *Despair*
6.67	3.33	**50**	6.67	6.67	10	16.67	0	c *Interest*
10	6.67	10	**53.33**	3.33	6.67	3.33	6.67	d *Irritation*
3.33	0	13.33	16.67	**53.33**	10	0	3.33	e *Joy*
6.67	13.33	6.67	0	6.67	**53.33**	13.33	0	f *Pleasure*
6.67	3.33	16.67	6.67	13.33	20	**33.33**	0	g *Pride*
3.33	6.67	3.33	20	0	13.33	6.67	**46.67**	h *Sadness*

6. Bimodal Classification

We have two modalities in the bimodal classification process. Their integration helps to provide better performance in emotion recognition as compared to that of unimodal systems. For our model, this suggests an integration of sign language gestures and facial expressions. These features bring in complementary information toward the recognition of emotions and, therefore, fusing both into one can help capture the nuances of both gesture dynamics and facial cues, which themselves may not carry a full context of emotion but together can improve the classification accuracy. We applied for this task a Bayesian classifier called BayesNet, in a similar environment as applied for unimodal classification. Features relevant to each modality were extracted separately and then processed for this purpose. At the feature-level fusion stage, those features are then integrated. This is due to the combined integration of features before classification, enabling the classifier to learn hand gesture relations coupled with facial expression relations simultaneously. The input data was normalized so as to ensure that consistency is maintained across the modalities. Feature discretization was then done using Kononenko's MDL criterion. Besides, we used WrapperSubsetEval to select the

best subset of features from the composite dataset as this reduces learning complexity and improves the classifier's performance. Cross-validation was used to check on the robustness of the classifier whereby we used a 10-fold cross-validation method to check the consistency of its performance on different subsets of data.

7. Discussion

In the systems that utilized only one type of input in the detection of emotions, the classifier built on gestures showed the best performance as it correctly classified the cases to 67.1%. The classifier built on facial expressions proved to have a lower accuracy of 48.3%, whereas a speech-based classifier reached 57.1%. This classifier is likely to be better because each expression in the dataset is clearly represented by some unique gesture, and this helps the classifier to identify and differentiate between the different expressions correctly. Participants were instructed to perform gestures tailored to convey different emotions, making it easier for the system to distinguish emotions from gestures. This explicit mapping likely made the emotion classification more straightforward than for facial or speech data, where emotional cues may be subtler or inconsistent.

Other research has shown strong results with body movement data as well. For example, Gunes and Piccardi reported a 90% recognition rate, and Bernhardt and Robinson achieved 81%. Although these results outperform the current study, it is important to note that these systems were designed to recognize fewer emotions (6 and 4 emotions, respectively) compared to the 8 emotions targeted here. Additionally, the current study used non-professional actors, which could have impacted the consistency and clarity of the gestures and, therefore, the system's performance.

Similarly, the system based on the expression of face did not match the high recognition rates found in some studies. This is likely due to the fact that our corpus was designed for multimodal recognition, which tends to perform worse when used in a unimodal context. Systems designed specifically for facial expression recognition, like those in the Cohn-Kanade DFAT-504 dataset, can achieve higher recognition rates, particularly when recognizing extreme expressions. Moreover, in this study, participants were not explicitly instructed on which facial expressions to display, leading to greater variability in facial cues for the same emotion. While this added to the naturalness of the dataset, it also made classification more difficult. For the speech-based classifier, comparison with other studies, such as Schuller et al. [25], highlights key differences. Schuller reported over 70% recognition accuracy using the Emo-DB database, but only around 54% with the DES database. One reason for this difference is the selection process used to create the speech corpus. In Emo-DB, samples were chosen based on perceptual clarity and naturalness, while in our study, no such selection or perceptual testing was performed, making it difficult to evaluate the effectiveness of the acted emotional expressions. Additionally, our study aimed to classify 8 emotions, compared to the 4 or 6 emotions in Schuller's work, adding complexity to the task.

Finally, recording conditions also played a role in the performance of the speech-based classifier. While the recordings were not excessively noisy, they were not conducted in studio conditions. The microphone setup, with a somewhat distant and non-directional mic, likely reduced the signal-to-noise ratio, negatively affecting the quality of the features extracted and, ultimately, the classification results.

8. Objective

The aim of this study is, therefore, to design a multimodal emotion recognition model for sign language by incorporating facial expressions as well as gesture-based features, in an attempt to improve the accuracy in the detection of emotions. The study compares the effectiveness of unimodal, bimodal, and multimodal systems toward demonstrating how combining facial expression and gesture data improves the overall recognition rate compared to the use of either modality alone. This research aims to support more accurate and naturalistic emotion recognition systems, especially for sign language communication.

9. Conclusion

It introduces a system that recognizes emotions by reading both facial expressions and gestures. The rationale is based on the idea that a combination of facial expressions and gestures creates a novel means of improving emotion recognition. It discusses how these two types of information can be combined and demonstrates how simultaneously using both results in a significantly improved accuracy compared with attention to one only. The system performs better as it uses multiple sources, similar to a human using different senses in order to understand the emotions. Moreover, this approach is especially helpful in cases where one source of information might miss or become unreliable, such as noisy environments or hard-to-capture data. Therefore, a system that can handle these challenges for practical use is essential.

This study underscores the system has to capture the timing and flow of emotional expressions. Facial expressions and gestures change with time, and the evolution incorporated into the changes formed is of much importance in ensuring the proper recognition of emotions. With these time-based features, the system gets an all-rounded, very realistic understanding of how emotions are presented, retaining valuable information about how emotions evolve over time. These dynamic features were often selected as the most critical accurate emotion recognition during the feature selection process. although this research was based on a limited dataset recorded in controlled conditions, it serves as a preliminary step toward fusing multiple synchronized modalities—an approach often discussed but less commonly implemented in emotion recognition research. Using a smaller dataset initially helps refine and optimize the data collection process, preparing for more extensive studies.

Future work will involve collecting larger multimodal datasets with a broader participant pool, ideally including spontaneous emotional expressions in naturalistic settings. This will introduce new challenges, such as dealing with occlusions, background noise, illumination changes, and head movements, which will need to be addressed comprehensively.

Additionally, future research will focus on developing advanced multimodal fusion techniques that better capture the relationships between features across different modalities, the correlation between visual and gestural data, and the informative contribution of each modality to the overall emotion recognition task.

References

1. *2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS 2015)* A New Data Glove Approach for Malaysian Sign Language Detection.
2. Bridging the Communication Gap: Artificial Agents Learning Sign Language through Imitation
3. Federico Tavella 1,2, Aphrodite Galata 2 and Angelo Cangelosi 1,2 (2024)
4. Ahmad Zaki Shukor, Muhammad Fahmi Miskon, Muhammad Herman Jamaluddin, Fariz bin
5. Ali@Ibrahim, Mohd Fareed Asyraf *, Mohd Bazli bin Bahar
6. A survey on recent advances in Sign Language Production Razieh Rastgoo a, ∗, Kourosh Kiani a, Sergio Escalera b, Vassilis Athitsos c, Mohammad Sabokrou d
7. Machine Learning with Applications 14 (2023) 100504 A survey on sign language literature Marie Alaghband a, b, ∗, Hamid Reza Maghroor b, Ivan Garibay b
8. Karg, M., Kuhnlenz, K., & Buss, M. (2010). Recognition of affect based on human body movement. *Pattern Recognition*, 43(3), 1052-1064.
9. Neural Sign Language Translation Necati Cihan Camgoz 1 , Simon Hadfield 1 , Oscar Koller 2 , Hermann Ney 2 , Richard Bowden 11 University of Surrey, {n.camgoz, s.hadfield, r.bowden}@surrey.ac.uk2 RWTH Aachen University, {koller, ney}@cs.rwth-aachen.de
10. Expert Systems with Applications 103 (2018) 159–183 Gesture recognition: A review focusing on sign language in a mobile contextDavi Hirafuji Neiva ∗, Cleber Zanchettin
11. Pantic, M., Valstar, M., Rademaker, R., & Maat, L. (2005). Web-based database for facial expression analysis. *IEEE International Conference on Multimedia and Expo*.
12. Benitez-Quiroz, C. F., Srinivasan, R., & Martinez, A. M. (2016). EmotioNet: An Accurate, Real-Time Algorithm for the Automatic Annotation of a Million Facial Expressions in the Wild. CVPR.
13. Sebe, N., Cohen, I., Gevers, T., & Huang, T. S. (2006). Multimodal approaches for emotion recognition: A survey. *Proceedings of SPIE-The International Society for Optical Engineering*.
14. Pantic, M., & Rothkrantz, L. J. (2003). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9), 1370-1390.
15. Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), 39-58.
16. Learning a deep network with spherical part model for 3D hand pose estimation Tzu-Yang Chen a , Pai-Wen Ting a , Min-Yu Wu a , Li-Chen Fu a,b,∗ Pattern Recognition 80 (2018) 1–20
17. Setia, S., Anjli, K., Bisht, U., Jyoti, Raj, D. Event Management System Using Spatial and Event Attribute Information. SN COMPUT. SCI. 6, 290 (2025). https://doi.org/10.1007/s42979-025-03781-0.
18. Naitik, D. Raj, D. K. Rajan, A. K. Gupta, A. K. Agrawal and K. R. Krishna, "Enhancing Toxic Comment Detection with BiLSTM-Based Deep Learning Model," *2024 International Conference on Information Science and Communications Technologies (ICISCT)*, Seoul, Korea, Republic of, 2024, pp. 206-211, doi: 10.1109/ICISCT64202.2024.10956568.
19. S. Singh, P. Prakash, G. Baghel, A. Singh, D. Raj and A. K. Agrawal, "Banana Crop Health: A Deep Learning-Based Model for Disease Detection and Classification," *2024 27th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Greater Noida, India, 2024, pp. 1-6, doi: 10.1109/WPMC63271.2024.10863138.
20. Adhikari, M.S., Gupta, R., Raj, D., Astya, R., Ather, D., Agrawal, A. (2025). Prevention of Attacks on Spanning Tree Protocol. In: Dutta, S., Bhattacharya, A., Shahnaz, C., Chakrabarti, S. (eds) *Cyber Intelligence and Information Retrieval. CIIR 2023. Lecture Notes in Networks and Systems*, vol 1139. Springer, Singapore. https://doi.org/10.1007/978-981-97-7603-0_24

21. D. Raj, A. K. Gupta and K. Rama Krishna, "Comparative Analysis of Different Approaches for Cyber Forensics," *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS),* Tashkent, Uzbekistan, 2024, pp. 42-47, doi: 10.1109/ICTACS62700.2024.10840964
22. Raj, D., Ather, D. & Sagar, A.K. Advancing Vehicular Ad-Hoc Network Solutions in Emerging Economies: A Comparative Analysis of V2V Protocols Through Simulation Studies. *SN COMPUT. SCI.* 5, 1077 (2024). https://doi.org/10.1007/s42979-024-03411-1
23. Bhardwaj, A., Sharma, A., Raj, D., Ather, D., Sagar, A. K., & Jain, V. (2025). Dynamic and Scalable Privacy-Preserving Group Data Sharing in Secure Cloud Computing. In N. Chaubey & N. Chaubey (Eds.), *Advanced Cyber Security Techniques for Data, Blockchain, IoT, and Network Protection* (pp. 89-122). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-9225-6.ch004
24. Singhal, R., Jain, V., & Raj, D. (2025). E-Health Transforming Healthcare Delivery With AI, Blockchain, and Cloud. In M. Lytras, A. Alkhaldi, & P. Ordóñez de Pablos (Eds.), *Harnessing AI, Blockchain, and Cloud Computing for Enhanced e-Government Services* (pp. 475-510). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-7678-2.ch015
25. Gupta, R., Adhikari, M. S., Raj, D., Jain, V., Sagar, A. K., & Ather, D. (2024). Blockchain in Web3.0. In K. Abhishek & C. Chakraborty (Eds.), *Blockchain-Based Solutions for Accessibility in Smart Cities* (pp. 171-204). IGI Global. https://doi.org/10.4018/979-8-3693-3402-7.ch007
26. Pranjal, Vaishnavi, Divyansh, Raj, D., Jain, V., Agarwal, A.K. (2024). "Adversarial Attacks on Neural Networks". In: Dutta, S., Bhattacharya, A., Shahnaz, C., Chakrabarti, S. (eds) *Cyber Intelligence and Information Retrieval*. CIIR 2023. Lecture Notes in Networks and Systems, vol 1025. Springer, Singapore. https://doi.org/10.1007/978-981-97-3594-5_34.
27. Gandhar A.; Gupta K.; Pandey A.K.; Raj D., "Fraud Detection Using Machine Learning and Deep Learning", *2024, SN Computer Scienc*e, Volume-5, Issue-5, DOI: 10.1007/s42979-024-02772-x
28. Prajapati A.; Gupta A.; Mishra S.; Raj D.; Singh M.K.; Goyal M.K., "An Exploration on Big Data Analytical Techniques: A Review", *2024, Proceedings of the 18th INDIAcom; 11th International Conference on Computing for Sustainable Global Development*, *INDIACom* 2024, pp: 123-128, DOI: 10.23919/INDIACom61295.2024.10498836
29. Chaudhary A.; Krishna K.C.; Shadik M.; Raj D., "Detection of Phishing Link Using Different Machine Learning Techniques", *2024, Lecture Notes in Networks and Systems,* Volume-896, PP: 63-77, DOI: 10.1007/978-981-99-9811-1_6
30. Rai S.; Upadhyay A.K.; Sharma D.; Raj D.; Gupta A.K.; Ather D., "Quantum cryptography-A modern approach", *2023, Journal of Discrete Mathematical Sciences and Cryptography*, Volume-26, Issue-7,PP: 1991-2006, DOI: 10.47974/JDMSC-1839
31. Raj D.; Sagar A.K., "Vehicular Ad-hoc Networks: A Review on Applications and Security", *2023, Communications in Computer and Information Science*, Volume-1921 CCIS, PP: 241-255, DOI: 10.1007/978-3-031-45124-9_19
32. Dhoundiyal S.; Arora A.; Mohakud S.; Patadia K.; Gupta A.K.; Raj D., "A Multilingual Text to Speech Engine Hindi-English: Hinglish", *2023, Proceedings of the 12th International Conference on System Modeling and Advancement in Research Trends, SMART 2023*, pp: 480-485, DOI: 10.1109/SMART59791.2023.10428607
33. Santhan A.; Tomar A.K.; Arora V.; Raj D., "Tank water flow automation", *2023, Artificial Intelligence, Blockchain, Computing and Security:* Volume 1, Volume-1, PP: 924-928, DOI: 10.1201/9781003393580-138
34. Chaudhary A.; Krishna K.C.; Shadik M.; Raj D., "A review on malicious link detection techniques", *2023, Artificial Intelligence, Blockchain, Computing and Security:* Volume 1, Volume-1, PP: 768-777, DOI: 10.1201/9781003393580-114

35. Ali S.A.; Roy N.R.; Raj D., "Software Defect Prediction using Machine Learning", *2023, Proceedings of the 17th INDIACom; 10th International Conference on Computing for Sustainable Global Development, INDIACom 2023*, pp: 639-642.
36. Borges, Tanya and Rai, Akash and Raj, Dharm and Ather, Danish and Gupta, Keshav, "Kidney Stone Detection using Ultrasound Images" (July 14, 2022). *Proceedings of the Advancement in Electronics & Communication Engineering 2022*, Available at http://dx.doi.org/10.2139/ssrn.4159208
37. Jain, Ashima and Sarkar, Arup and Ather, Danish and Raj, Dharm, "Temperature Based Automatic Fan Speed Control System using Arduino" (July 14, 2022). *Proceedings of the Advancement in Electronics & Communication Engineering 2022*, Available at SSRN: http://dx.doi.org/10.2139/ssrn.4159188 .
38. Challa, Neha and Baishya, Kriti and Rohatgi, Vinayak and Gupta, Keshav and Ather, Danish and Raj, Dharm, Recent Advances in Sign Language Detection: A Brief Survey (July 14, 2022). *Proceedings of the Advancement in Electronics & Communication Engineering 2022*, Available at http://dx.doi.org/10.2139/ssrn.4157565
39. Malhotra, Chiranjeev and, Devanshu and Sharma, Sourav and Arquam, Md. and Maini, Tarun and Raj, Dharm, "Complete Medical Solutions With InstaMedi" (July 14, 2022). *Proceedings of the Advancement in Electronics & Communication Engineering 2022*, Available at http://dx.doi.org/10.2139/ssrn.4159204
40. Agarwal A.K.; Ather D.; Astya R.; Parygin D.; Garg A.; Raj D., "Analysis of Environmental Factors for Smart Farming: An Internet of Things Based Approach", *2021, Proceedings of the 10th International Conference on System Modeling and Advancement in Research Trends, SMART 2021*, pp: 210-214, DOI: 10.1109/SMART52563.2021.9676305
41. Ojha R.P.; Raj D.; Srivastava P.K.; Sanyal G., "Gaussian tendencies in data flow in communication links", *2018, Advances in Intelligent Systems and Computing*, Volume-729, PP: 499-505, DOI: 10.1007/978-981-10-8536-9_48
42. Srivastava M.; Singh H.M.; Gupta M.; Raj D., "Digital watermarking using spatial domain and triple des", *2016, Proceedings of the 10th INDIACom; 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016*, pp: 3031–3035.
43. Kumar S.; Raj D., "A contemporary approach to hybrid expert systems: Case base reasoning", 2010, *2010 International Conference on Computer and Communication Technology*, ICCCT-2010, pp: 736-740, DOI: 10.1109/ICCCT.2010.5640376
44. Raj D.; Tripathi R.C., "Method for generating 3-dimensional wireframe model from different 2-dimensional drawings", 2010, NISS2010 - *4th International Conference on New Trends in Information Science and Service Science*, pp: 313-318.