Amyotrophic lateral sclerosis Disease Detection Through Facial Expression Analysis

Sri Varshan P
Department of Computer Science and Engineering
Amrita School of Computing, Amrita Vishwa Vidyapeetham
Chennai campus
immadevarshan7@gmail.com

Harigaran R
Department of Computer Science and Engineering
Amrita School of Computing, Amrita Vishwa Vidyapeetham
Chennai Campus
harigaranramesh23@gmail.com

Vamsi Krishna V
Department of Computer Science and Engineering
Amrita School of Computing, Amrita Vishwa Vidyapeetham
Chennai Campus
v_vamsikrishna@ch.amrita.edu

2025 International Conference on Ambient Intelligence in Health Care (ICAIHC) | 979-8-3315-3025-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICAIHC64101.2025.10956283

***Abstract*—Identifying Amyotrophic Lateral Sclerosis (ALS) early is crucial for initiating treatment, improving prognosis, and enhancing the quality of life for affected individuals. However, early diagnosis and symptom detection can be challenging. A more accessible and cost-effective solution involves computational analysis of the patient's facial expressions. ALS patients exhibit distinct facial muscle movements during specific actions, such as mouth opening, compared to healthy individuals. This paper proposes using deep learning-based facial expression analysis to detect ALS. The model was trained to recognize ALS symptoms through facial expression data and evaluated on the NeuroFace dataset. Experimental results show that the model achieved a high precision of 0.94, a recall of 0.89, and an F1-score of 0.92 for the ALS class. The overall accuracy of the model was 92%, demonstrating its effectiveness in distinguishing between ALS and healthy individuals, thus contributing to early ALS detection through non-invasive methods.**

***Index Terms*—Amyotrophic lateral sclerosis, microexpression, OpenCV, Facial Landmark, Mediapipe.**

I. Introduction

Neurological disorders, such as Amyotrophic Lateral Sclerosis (ALS), have profound impacts on patients' quality of life and present significant challenges in early diagnosis and monitoring. Traditional diagnostic methods often involve invasive procedures or expensive imaging techniques, limiting their accessibility and frequency of use. In recent years, the intersection of computer vision, machine learning, and healthcare has opened new avenues for non-invasive diagnostic tools. This research focuses on a novel approach to neurological disease detection through the analysis of facial expressions. The annual incidence rate of ALS is reported to be between 1.5 and 2.5 per 100,000, with a prevalence ranging from 2 to 7 per 100,000 [1]. In North America, over 30,000 individuals are currently living with ALS, and more than 5,600 new cases are diagnosed each year [2]. The human face is a rich source of information, capable of conveying complex emotional and physiological states. Subtle changes in facial movements and expressions can potentially indicate the presence and progression of neurological disorders. By leveraging advanced computer vision techniques and deep learning algorithms, we propose a system that can detect and analyze micro-expressions, facial landmarks, and temporal changes in facial movements to identify markers of neurological diseases, with a particular focus on ALS. Our approach utilizes a multi-modal system that combines: Facial landmark detection using MediaPipe, Micro-expression analysis through optical flow calculations, Feature extraction and dimensionality reduction using Principal Component Analysis (PCA), A sophisticated multi-input neural network for classification. This research aims to develop a cost-effective, non-invasive diagnostic tool that can be widely deployed in various healthcare settings. By analyzing video data of patients performing specific oro-facial gestures, our system seeks to identify subtle indicators of neurological disorders that may not be immediately apparent to the human eye. Beyond the significant personal and social burdens, ALS is also an economically costly disease. Recent estimates place the nationwide cost of ALS in the U.S. at $1.023 billion, with annual expenses per individual averaging $64,000 in the U.S. and $32,000 in Canada [3]. The potential impact of this research extends beyond the initial diagnosis. Such a system could provide a means for continuous monitoring of disease progression, enabling more personalized treatment plans and earlier interventions. Moreover, it could serve as a valuable tool for telemedicine applications, allowing for remote assessment and follow-up of patients with suspected or confirmed neurological disorders. In this paper, we present the design, implementation, and evaluation of our proposed system, utilizing the Toronto NeuroFace Dataset [4] as our benchmark. We discuss the challenges encountered, the solutions developed, and the performance of our model compared to existing systems.

Finally, we explore the implications of our findings and suggest directions for future research in this promising field of non-invasive neurological disease detection.

II. RELATED WORKS

Expression in the face (FEs) has a vital role in analysis of behavior and face-to-face interactions. Mehrabian et al. [5] mentioned that, for impactive verbal communication, factors like voice tone and words contribute a total importance of 45% only whereas the remaining 55% depends on body language which includes facial expressions. Assessing and rehabilitating facial impairments through automated video-based techniques has been a major area of research[6-12] Parkinson's disease [6-9] and the after-effects of stroke[8-10] have been studied in the majority of papers up to date. In [6], the Authors performed facial alignment through the Supervised Descent Method(SDM), and lip movements were examined using Microsoft Kinect. The findings revealed that patients with Parkinson's showed decreased acceleration and velocity in their lip movements during syllable repetition tasks, compared to healthy control (HC) subjects. These analysis were linked to the diagnosis of hypokinetic dysarthria, a hallmark of Parkinson's. Several researches [7-9] explored hypokinesia and hypomimia in Parkinson's disorder, characterized by a diminished ability to express facial emotions. In [7], facial landmarks were analyzed in RGB videos using SDM, and 20 geometric features [16] were extracted from neutral expressions, as well as expressions that are imitated. The study found that the difference between the features of expressive and the neutral face was decreased in PD patients. Additional studies [8,9] analyzed facial actions and expressions in PD by identifying Action Units (AUs) [17]. Specifically, [9] employed geometric and appearance-based features along with Support Vector Machines (SVM) to detect AUs, demonstrating reduced expressivity in Parkinson's and its correlation with the seriousness of the disease. Similarly, several approaches have been developed to analyze facial asymmetries resulting from stroke or any damage to facial nerves[10-12]. In [10], 3D facial scans were used to measure asymmetry by calculating the Euclidean distance between corresponding points on the original and mirrored 3D facial structures. A greater distance indicated more severe facial asymmetry, detected in stroke patients during speech tasks. Facial movements have been analyzed as part of rehabilitation efforts. A Kinect-based automated system described in [14] provided visual feedback for facial paralysis therapy. This system utilized geometric and surface curvature features derived from 12 facial regions to classify designated facial exercises using a random forest model. These findings demonstrate the effectiveness of depth sensors and face-tracking technologies in diagnosing and managing facial impairments related to neurological conditions.

III. METHODOLOGY

A. Dataset Collection

This paper uses the Toronto neuroface dataset[4], a publicly available dataset collected by recording individuals with neurological disorders' orofacial gestures. The dataset is divided into three labels: Amyotrophic Lateral Sclerosis(ALS): Collected from individuals with orofacial impairment due to ALS. Stroke: Collected from participants whose orofacial gestures have been affected post-stroke. Healthy Controls: Data of healthy individuals. The dataset contains videos of these individuals performing various orofacial tasks which includes speech exercises and non-speech movements. The videos captures participants performing tasks such as repeating the sentence "Buy Bobby a Puppy", Rapid repetition of the syllable 'pa' and the sequence 'pataka' as quickly as possible in a single breath, Non-speech movements, such as lip puckering (pretending to blow a candle or kiss) and maximum jaw opening, Expressive movements, such as smiling and raising eyebrows. This paper uses ALS and healthy control labels alone to perform facial expression analysis, as illustrated in Figure 1, which outlines the overall proposed methodology.

B. Data Preprocessing

The frames extracted from the videos are organized into separate folders for each video and sequentially ordered to maintain the temporal order of each video. Extracted frames from the dataset are already resized to a standard dimension of 640x480. The pixel values are normalized between 0 and 1 to reduce variance during training.

C. Feature Extraction

In this paper two distinct feature extraction techniques are used to detect signs of Amyotrophic lateral sclerosis in individuals: Microexpression features and facial landmark features. Microexpression features focus on the subtle facial movements that are difficult to capture with the naked eye. Whereas landmark features provide information about the relative positions of key facial points such as eyes, mouth, nose, etc, focusing on the large-scale changes of the face during expression. Combining the intricate microexpression features and abstract landmark features can help the model capture information about small-scale muscle deviation and also high-level structural changes related to ALS.

*1) **Microexpression Feature Extraction**:* For extracting microexpression features, optical flow is calculated between consecutive frames using the Farneback method. OpenCV library's inbuilt function `calcOpticalFlowFarneback` [18] is used. First, each video frame is converted to grayscale to reduce computational complexity and focus on intensity changes. The Farneback method estimates optical flow by approximating pixel motion using polynomial expansions across multiple pyramid levels.

$$I_1(x, y) = I_2(x + u, y + v) \tag{1}$$

$$\frac{\partial I_1}{\partial x}u + \frac{\partial I_1}{\partial y}v = -\frac{\partial I_1}{\partial t} \tag{2}$$

$$M = \sqrt{u^2 + v^2} \tag{3}$$

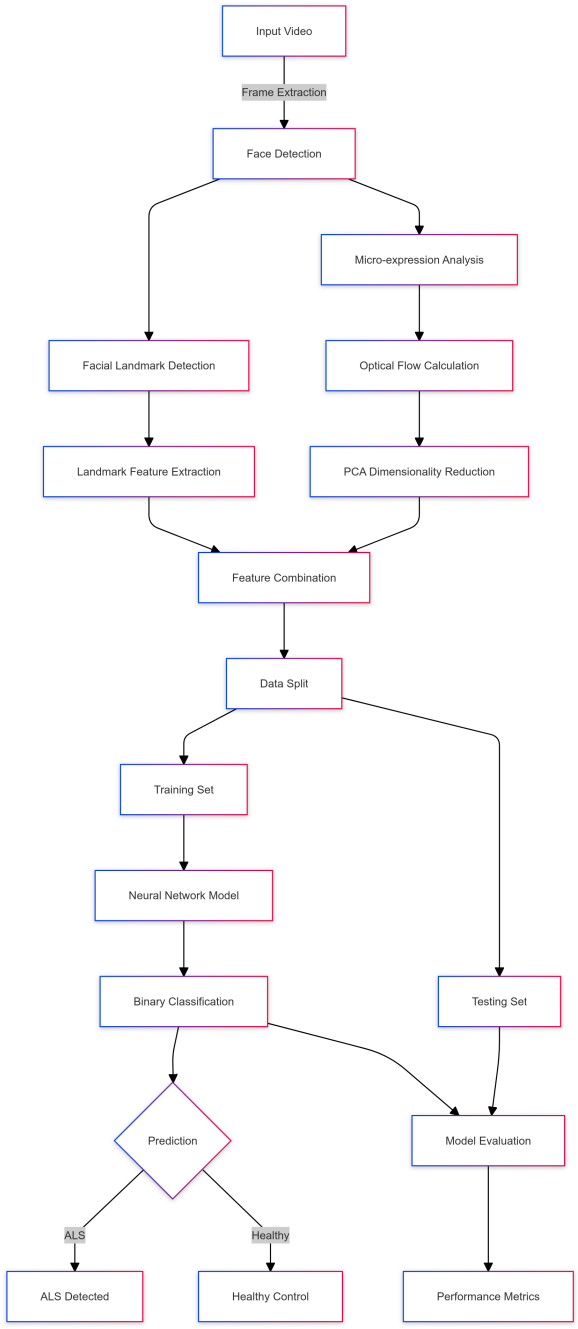

Fig. 1. Workflow of the proposed methodology

$$\theta = \text{atan2}(v, u) \tag{4}$$

the motion vectors at each pixel (x,y) are calculated by solving for the displacement vector u,v(1,2), which shows the optical flow constraint formula used to calculate the motion vector of each pixel value. The resulting motion vectors are then converted from Cartesian coordinates to polar coordinates using `cv2.cartToPolar` [19], yielding the magnitude (speed) and angle (direction) of motion at each pixel. u and v are the horizontal and vertical components of the motion vector at pixel (x,y)(x, y)(x,y) (3) shows the mathematical formula for calculating the magnitude and direction of the motion vector. atan2 is the arctangent function, which calculates the angle of the vector from the horizontal axis.(4)

The motion magnitude is flattened into a 1D array for each frame, forming the microexpression features that describe the fine-grained facial movement patterns.

*2) **Landmark Feature Extraction**:* This paper uses Mediapipe [20], an open-source library developed by Google, for extracting facial landmark features. Mediapipe's face-mesh model extracts 468 facial landmarks from the human face. The dense landmark features help the model capture the patterns in facial movements with high precision. Each video frame is processed by converting it to RGB format. Mediapipe's Face Mesh model is used to detect landmarks, returning the x and y coordinates of key facial points. The coordinates of these landmarks are then flattened into a one-dimensional array, which serves as a feature vector representing the facial structure in the frame.

*3) **Dimensionality Reduction**:* As the feature size of microexpressions is relatively larger than that of landmark features due to its dense nature, the dimension of microexpression features is reduced through Principal Component Analysis (PCA) while preserving essential information. Reducing the dimension before combining the features will prevent the model from having a bias towards microexpression features when making decisions.

D. Model Training

The classification model is a neural network with two input branches to handle both types of features (landmark-based and microexpression-based). The structure of the model is as follows:

- **Landmark Input Branch:** The facial landmark features are passed through a fully connected dense layer with 128 units and a ReLU activation function. This is followed by a dropout layer with a rate of 0.5 to reduce overfitting. A second dense layer with 64 units and ReLU activation refines the landmark features.
- **Microexpression Input Branch:** The microexpression (optical flow) features are passed through a dense layer with 128 units and ReLU activation, followed by a dropout layer with a rate of 0.5 to combat overfitting. Another dense layer with 64 units is applied to further refine the features.
- **Feature Concatenation:** The outputs from both branches are concatenated and passed through additional dense layers for joint feature learning. The final layer is a single neuron with a sigmoid activation function, providing the binary classification result (ALS or Healthy). The network is trained using binary cross-entropy as the loss function and the Adam optimizer for efficient learning.

The proposed neural network model employs a dual-branch architecture to process microexpression and landmark features independently. Each branch consists of two fully connected

layers with ReLU activation and dropout for regularization, which extract high-level representations from the input features. The branches are concatenated and further processed through a fully connected layer before producing a binary classification output using a sigmoid activation function. The model is trained with the Adam optimizer and binary cross-entropy loss, achieving effective integration of spatial and temporal features for ALS detection.

IV. RESULTS AND DISCUSSION

We have trained a model on a dataset to recognize whether an individual has ALS symptoms or not through facial expression analysis in this paper. Precision for the healthy class is 0.90, that is, 90% of all people predicted to be healthy are indeed healthy. The recall is even more impressive, at 0.94, indicating that the model correctly identifies 94% of all healthy cases. The F1-score balancing both precision and recall sits at an impressive 0.92, showing good performance for this class. The overall performance metrics, including accuracy, macro average, and weighted average, as detailed in Table I, illustrate the robustness of the proposed model. The ALS Class: For

Class	Precision	Recall	F1-Score	Support
Healthy	0.90	0.94	0.92	169
ALS	0.94	0.89	0.92	169
Accuracy			**0.92**	338
Macro avg	0.92	0.92	0.92	338
Weighted avg	0.92	0.92	0.92	338

TABLE I
PERFORMANCE METRICS FOR HEALTHY AND ALS CLASSES USING THE PROPOSED MODEL.

ALS, the precision was 0.94, meaning that the model correctly classifies the ALS participants who it predicts to have it with 94% accuracy. On the other hand, the recall was a little lower at 0.89, meaning that 11% of the actual ALS cases were missed. Nonetheless, the F1-score for ALS is still quite good at 0.92; this means that the model is very close to being as right as possible while not picking too many false positives.

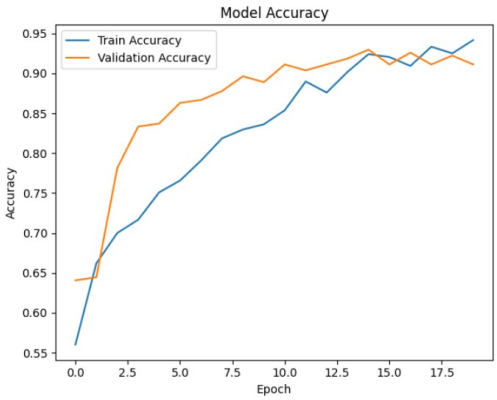

Fig. 2. Accuracy curves for training and validation sets over 20 epochs

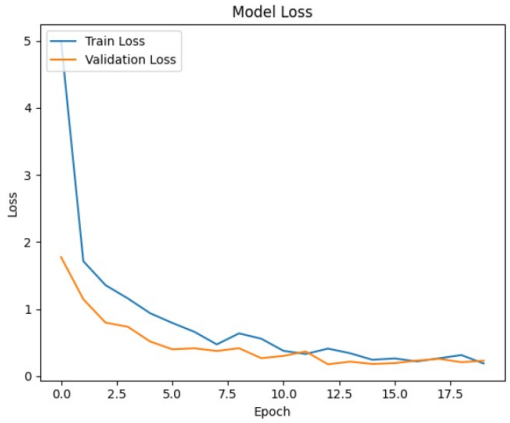

Fig. 3. Loss curves for training and validation sets over 20 epochs

The model accuracy on all predictions made in the overall test set is 0.92; that is, accuracy overall is 92%. In addition, the precision, recall, and F1-score macro average metrics are also 0.92, showing the weightage of both ALS and Healthy classes under the model's robust performance. Figures 2 and 3 are the accuracy and loss curves over the 28 training epochs, Training and validation accuracy increase monotonically while validation stabilizes around 92%. Validation sometimes being higher than training accuracy also shows that the model generalizes well and is not overfitting to the training data. The loss curves indicate a sharp decline within the first few epochs for both training and validation sets. Both start stabilizing around about 20 epochs. Low final values of loss show that the model is effectively minimizing error across both datasets, which again confirms that the model is indeed good at generalization to unseen data.

V. CONCLUSION

This paper demonstrates a new method of discovering ALS through facial expression analysis with the Toronto NeuroFace dataset. With the application of both landmark features and microexpression analysis, such as the computation of optical flow, the movements orofacial that tend to be related to ALS could be collected. Processing these combined features through a neural network model helps in performing the binary classification between ALS and controls who are healthy. We balanced our model with an accuracy of 92% and this is reflected through precision, recall, and F1-score for both ALS and Healthy classes. Results are promising for distinguishing patients with ALS from healthy subjects using automated detection method based on facial expression analysis. This workflow composed of feature extraction and dimensionality reduction followed by the classification process provides an effective and scalable pipeline that could easily be extended to larger datasets and diverse populations.

With promising results, such an approach could offer a helpful adjunct tool for ALS early diagnosis by clinicians with increased speed and accuracy in detecting ALS.

VI. ACKNOWLEDGMENT

Portions of the research in this paper use the Toronto NeuroFace Dataset for training and validating the machine learning model developed for facial expression analysis. the research in this paper uses the Toronto NeuroFace Dataset collected by Dr. Yana Yunusova and the Vocal Tract Visualization and Bulbar Function Lab teams at UHN-Toronto Rehabilitation Institute and Sunnybrook Research Institute respectively, financially supported by the Michael J. Fox Foundation, NIH-NIDCD, Natural Sciences and Engineering Research Council, Heart and Stroke Foundation Canadian Partnership for Stroke Recovery and AGE-WELL NCE

REFERENCES

[1] Brown, R. H., & Al-Chalabi, A. (2017). Amyotrophic Lateral Sclerosis. The New England journal of medicine, 377(2), 162–172. https://doi.org/10.1056/NEJMra1603471

[2] http://www.alsa.org [Accessed: October 5, 2024]

[3] Zahraa Hassan Shehata, Hoda Rabea, Rasha El Sherif, Mohamed E. Abdelrahim, Dalia M. Dawoud, Estimating Societal Cost of Illness and Patients' Quality of Life of Duchenne Muscular Dystrophy in Egypt, Value in Health Regional Issues, Volume 33, 2023, Pages 10-16, ISSN 2212-1099, https://doi.org/10.1016/j.vhri.2022.08.006.

[4] Bandini, A., Rezaei, S., Guarin, D. L., Kulkarni, M., Lim, D., Boulos, M. I., Zinman, L., Yunusova, Y., & Taati, B. (2021). A New Dataset for Facial Motion Analysis in Individuals With Neurological Disorders. IEEE journal of biomedical and health informatics, 25(4), 1111–1119. https://doi.org/10.1109/JBHI.2020.3019242

[5] Mehrabian, A. (1968). Some referents and measures of nonverbal behavior. Behavior Research Methods & Instrumentation, 1, 203-207.

[6] Bandini, A., Orlandi, S., Giovannelli, F., Felici, A., Cincotta, M., Clemente, D., Vanni, P., Zaccara, G., & Manfredi, C. (2016). Markerless Analysis of Articulatory Movements in Patients With Parkinson's Disease. Journal of Voice, 30(6), 766.e1-766.e11. https://doi.org/10.1016/j.jvoice.2015.10.01

[7] Bandini, A., Orlandi, S., Escalante, H. J., Giovannelli, F., Cincotta, M., Reyes-Garcia, C. A., Vanni, P., Zaccara, G., & Manfredi, C. (2017). Analysis of facial expressions in parkinson's disease through video-based automatic methods. Journal of Neuroscience Methods, 281, 7-20. https://doi.org/10.1016/j.jneumeth.2017.02.006

[8] Vinokurov, N., Arkadir, D., Linetsky, E., Bergman, H., Weinshall, D. (2016). Quantifying Hypomimia in Parkinson Patients Using a Depth Camera. In: Serino, S., Matic, A., Giakoumis, D., Lopez, G., Cipresso, P. (eds) Pervasive Computing Paradigms for Mental Health. MindCare 2015. Communications in Computer and Information Science, vol 604. Springer, Cham. https://doi.org/10.1007/978-3-319-32270-4_7

[9] Wu, P., Gonzalez, I., Patsis, G., Jiang, D., Sahli, H., Kerckhofs, E., & Vandekerckhove, M. (2014). Objectifying facial expressivity assessment of Parkinson's patients: preliminary study. Computational and mathematical methods in medicine, 2014, 427826. https://doi.org/10.1155/2014/427826

[10] W. Quan, B. J. Matuszewski and L. -K. Shark, "Facial asymmetry analysis based on 3-D dynamic scans," 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, Korea (South), 2012, pp. 2676-2681, doi: 10.1109/ICSMC.2012.6378151.

[11] J. R. Delannoy and T. E. Ward, "A preliminary investigation into the use of machine vision techniques for automating facial paralysis rehabilitation therapy," IET Irish Signals and Systems Conference (ISSC 2010), Cork, 2010, pp. 228-232, doi: 10.1049/cp.2010.0517.

[12] M. Schätz, A. Procházka, O. Ťupa, O. Vyšata and V. Sedlák, "Face movement analysis with MS Kinect," 2016 International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM), Reggio Calabria, Italy, 2016, pp. 1-5, doi: 10.1109/IWCIM.2016.7801179.

[13] Dittmar, Cornelia & Olgay, B.S. & Denzler, Joachim & Gross, Horst-Michael. (2013). Automated classification of therapeutic face exercises using the kinect. VISAPP 2013 - Proceedings of the International Conference on Computer Vision Theory and Applications. 1. 556-565.

[14] C. Dittmar, J. Denzler and H. -M. Gross, "A Feedback Estimation Approach for Therapeutic Facial Training," 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 2017, pp. 141-148, doi: 10.1109/FG.2017.26.

[15] X. Xiong and F. De la Torre, "Supervised Descent Method and Its Applications to Face Alignment," 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 2013, pp. 532-539, doi: 10.1109/CVPR.2013.75.

[16] M. Soleymani, J. Lichtenauer, T. Pun and M. Pantic, "A Multimodal Database for Affect Recognition and Implicit Tagging," in IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 42-55, Jan.-March 2012, doi: 10.1109/T-AFFC.2011.25.

[17] Ekman, P., & Friesen, W. V. (1978). Manual for the Facial Action Code. Palo Alto, CA: Consulting Psychologist Press.

[18] Object tracking (no date) OpenCV. Available at: https://docs.opencv.org/3.4/dc/d6b/group__video__track.html (Accessed: 10 October 2024).

[19] Cv2.CartToPolar Method (no date). Available at: https://shimat.github.io/opencvsharp_docs/html/d8325d06-b4c4-2e2e-9eec-dee08f6e8098.htm (Accessed: 10 October 2024).

[20] MediaPipe: A Framework for Perceiving and Processing Reality. (n.d.). Available at: https://research.google/pubs/mediapipe-a-framework-for-perceiving-and-processing-reality/ (Accessed: 10 October 2024).