**AMERICAN SIGN LANGUAGE DETECTION USING LSTM**

Debopriya Dutta Sumit Panja

*Department of Computer Science & Technology Institute of Engineering & Management (IEM), Kolkata*

*West Bengal, India*

[debopriyadutta012@gmail.com](mailto:debopriyadutta012@gmail.com)

[sumit123panja@gmail.com](mailto:debopriyadutta012@gmail.com)

**Abstract—** This work explores the potential of Long Short-Term Memory (LSTM) networks in the development of an automated American Sign Language (ASL) recognition system. In close consultation with experts in the domain, this project seeks to bridge communication gaps for individuals with hearing impairments. The LSTM-based system has been meticulously trained on a comprehensive dataset, demonstrating exceptional accuracy in interpreting a broad range of ASL gestures. This research highlights the power of machine learning in creating solutions that empower the hearing-impaired community. Its adaptability and proficiency, resulting from training on a rigorously curated dataset, showcase its capacity to recognize diverse sign language expressions. This work signifies a substantial advancement in assistive technologies, promoting a more inclusive and accessible world for all.

***Keywords***—***Long Short-Term Memory (LSTM) networks, American Sign Language (ASL) Detection, Sign language expressions, Machine Learning, Python, Machine learning, TensorFlow, Media pipe Holistic model***

**INTRODUCTION**

Sign language empowers communication for countless deaf and hard-of-hearing individuals worldwide. Yet, a communication divide often exists when interacting with those who do not understand this gestural language. Our project confronts this challenge by developing an innovative real-time sign language detection system. By integrating leading-edge technologies – Media Pipe for accurate hand landmark identification, LSTM RNNs (powered by TensorFlow) for analyzing sequential patterns, and OpenCV for computer vision functions – we strive to construct a transformative communication bridge. This system will skillfully recognize sign language gestures in real-time, rendering them into accessible formats like on-screen text. Our endeavor promises to revolutionize communication accessibility, promoting a more inclusive and connected world for deaf, hard-of-hearing, and hearing individuals alike.

The existing communication gap limits social interactions, access to information, and full participation in daily life for those relying on sign language. While sign language interpreters are invaluable, their availability is often limited and resource-dependent. Moreover, text-based communication alone cannot fully capture the visual nuances of sign language. Our research addresses the urgent need for a more intuitive and widely deployable solution to facilitate effortless communication.

The proposed sign language detection system acts as a perceptive interpreter. Employing OpenCV for live video capture and Media Pipe for meticulous hand landmark analysis within each frame, it feeds data into a sophisticated LSTM RNN model. This carefully trained model interprets the temporal patterns of sign language gestures, translating them into understandable text or other desired outputs. Our system transcends basic communication needs, holding the potential to empower deaf students within educational settings and augment accessibility across video conferencing platforms. An emphasis on user-centric design will ensure ease of use for individuals with varying technical backgrounds.

Our primary focus is serving the deaf and hard-of-hearing community who depend on sign language. Simultaneously, the system offers valuable benefits to secondary audiences, such as educators of deaf students and professionals who interact with these communities. Future refinements in pose and gesture recognition will drive greater accuracy, speed, and scalability. Harnessing the latest machine learning advancements and AI-driven personalization, Media Pipe is positioned to spearhead these advancements, propelling innovation in interactive technology across diverse areas.

**AREA OF WORK**

This project demonstrates a focused and deliberate approach, concentrating specifically on the detection of American Sign Language (ASL). This targeted emphasis suggests a potential application in accessibility enhancement or furthering human-computer interactions. The primary objective appears to be the automation of ASL gesture interpretation, potentially enabling real-time communication or assistive technologies for the deaf and hard-of-hearing. The use of LSTM networks underscores the employment of sophisticated machine learning techniques to achieve robust sign language recognition. Python, as the chosen programming language, reinforces a commitment to adaptability and consistency throughout the development process. Collectively, these aspects outline a project with a well-defined purpose within the broader field of sign language recognition. The work holds the potential to create significant advancements in bridging communication gaps for those who rely on ASL.

**APPLICATIONS**

The potential applications of sign language detection extend across a wide range of fields, offering transformative benefits in accessibility, communication, and interaction for those who rely on sign language. Let's examine these applications in detail:

1. **Communication Access:** Real-time interpretation of sign language gestures is a cornerstone application. This enables fluid communication between deaf or hard-of-hearing individuals and the hearing population in diverse settings, bridging long-standing communication gaps.
2. **Educational Support:** Sign language detection provides invaluable assistance to deaf students. Translating sign language into text or spoken language promotes equitable access to learning materials and fosters a truly inclusive classroom environment.
3. **Assistive Technology:** When integrated into wearable devices or smartphone applications, sign language detection offers a powerful communication tool for individuals with hearing impairments. This empowers users to interact with their environment, access information, and maintain greater independence.
4. **Accessibility in Public Services:** Critical areas like healthcare, emergency services, and public facilities see a significant boost in accessibility through sign language detection technology. This ensures that deaf individuals receive the same quality of service and remain fully informed.
5. **Employment Opportunities:** By removing communication roadblocks within workplaces, sign language detection champions a more inclusive and diverse work environment. This supports the hiring, training, and productive collaboration of deaf employees across various industries, leveling the playing field for career advancement.
6. **Cultural Preservation:** Sign language detection serves as a tool for safeguarding and promoting sign language as a fundamental pillar of deaf culture. It aids linguistic research, educational initiatives, and advocacy efforts, ultimately preserving the unique beauty and diversity of sign languages around the world.

**EXTRACTING FEATURES**

Our proposed methodology harnesses the strengths of the Media Pipe library for comprehensive feature extraction. A holistic model within the library locates essential landmarks across the face, hands, and body, providing rich spatial data about human form and movement within each video frame. We extract coordinates and visibility information for these landmarks, formatting the data for optimal use in machine learning algorithms.

Feature vectors are constructed by flattening the coordinates of pose, right hand, left hand, and face landmarks into one-dimensional arrays for each frame. This effectively captures spatial relationships across a sequence of frames. We utilize deep learning methods to derive high-level features from these key points. In particular, Long Short-Term Memory (LSTM) networks or other suitable recurrent neural networks (RNNs) model the temporal dependencies between successive frames. This empowers the system to discern patterns and gestures over time, crucial for accurate sign language interpretation.

To further augment our approach, we investigate optical flow algorithms like Farneback and Lucas-Kanade. These algorithms capture motion information inherent in video sequences, supplementing the spatial data extracted from key points. This enriches the system's understanding of the dynamic nature of sign language gestures.

In summary, our feature extraction methodology employs a multi-pronged approach, integrating spatial and temporal information. It leverages both the power of deep learning and the insights from traditional computer vision techniques for robust ASL detection and translation.

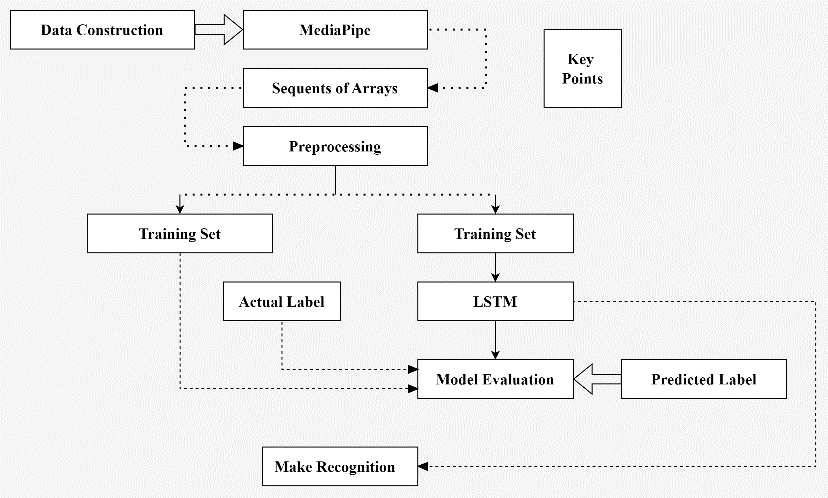
**METHODS**

This project endeavors to develop an American Sign Language (ASL) gesture recognition system integrating the strengths of Media Pipe and LSTM networks. Underscoring the importance of gesture recognition in empowering communication for those with hearing impairments, our objective is to facilitate seamless interpretation of ASL. Our methodology emphasizes a high-resolution webcam for capturing clear video data, adjustable lighting for optimal illumination, and a potentially uniform background to streamline image processing.

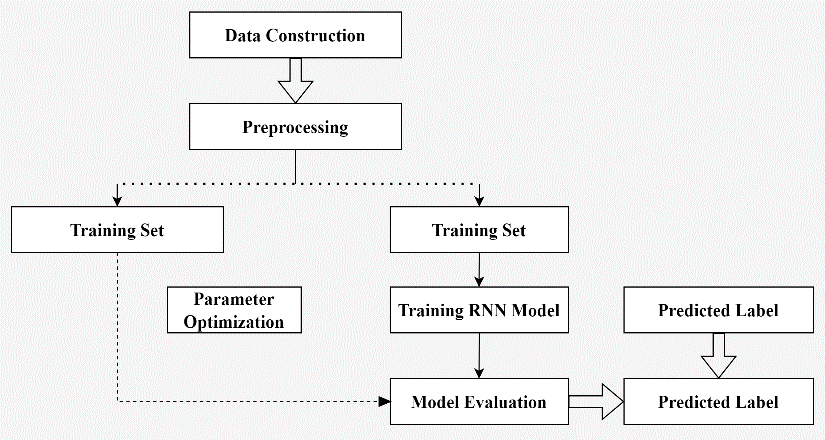
We outline a meticulous data collection and preprocessing approach. The acquisition process encompasses hardware setup and recruitment of participants possessing diverse hand shapes and sizes. To maintain consistency and high data quality, we'll provide clear recording instructions and ensure a fixed distance between participants and the camera. Data labelling and preprocessing involve extracting key point information using Media Pipe’s pose estimation capabilities. This data will be normalized and segmented into sequences representing individual gestures. For the model development phase, we'll design an architecture utilizing LSTM layers to analyze temporal dependencies between key points. The model will then undergo rigorous training and evaluation to optimize its performance.

Real-time gesture recognition will necessitate continuous processing of live video frames. Extracted key points will be transformed into sequences formatted for compatibility with the trained model, enabling real-time gesture prediction. The visualization and output stage will involve overlaying the predicted gesture label onto the live video and initiating pre-defined actions triggered by recognized gestures.

In our conclusion, we will summarize pivotal findings, delve into limitations, and propose potential avenues for enhancing the system's accuracy, real-time responsiveness, and robustness. To create a thoroughly robust and ethically sound ASL gesture recognition system, we will address crucial concerns such as data augmentation, background subtraction, optimizing real-time performance, error handling, user fairness, and ethical considerations.



*Figure 1. Block diagram of the proposed model for LSTM*

**

*Figure 2. Block diagram of the proposed model for RNN*

Figure 1. and Figure 2. can be used to show how the proposed method works. The dataset for the first proposed model diagram was obtained from the media pipe holistic pose estimation. Essential points are collected by following the process described in Figure 1. Using a holistic media pipe library, points are accumulated from hand, face, and pose landmarks. After preprocessing several frames, the arrays are transferred into LSTM.

**LSTM ARCHITECTURE**

Our project leverages the power of Media Pipe’s Holistic model, a Google-developed open-source framework. This model integrates pose, hand, and face detection for comprehensive body tracking within a video frame. It precisely locates 512 key points: 468 for facial landmarks, 33 for pose estimation, and 21 for each hand. Media Pipe Holistic provides rich spatial data, including x, y, and z coordinates, along with visibility scores indicating detection confidence.

After Media Pipe Holistic identifies relevant body parts, we extract the essential key point data for action classification. This data, including spatial coordinates and visibility scores, forms the basis for training our gesture recognition model. We'll establish a predetermined set of distinct actions, each defined by specific body movements, hand gestures, and facial expressions. This set determines the LSTM model's output layer size, allowing for accurate action classification based on temporal patterns across video frames.

The model architecture centers around stacked LSTM layers, specifically designed to analyze long-term dependencies within sequential data. This is vital for interpreting actions over time. The model incorporates carefully crafted input layers to process key point sequences and dense layers for feature extraction. We'll employ the Adam optimizer and categorical crossentropy loss during training to ensure the model accurately predicts actions. The LSTM model harnesses temporal dependencies in real-time video input, using a thresholding mechanism for reliable and robust gesture recognition.

**REAL-TIME SIGN LANGUAGE DETECTION**

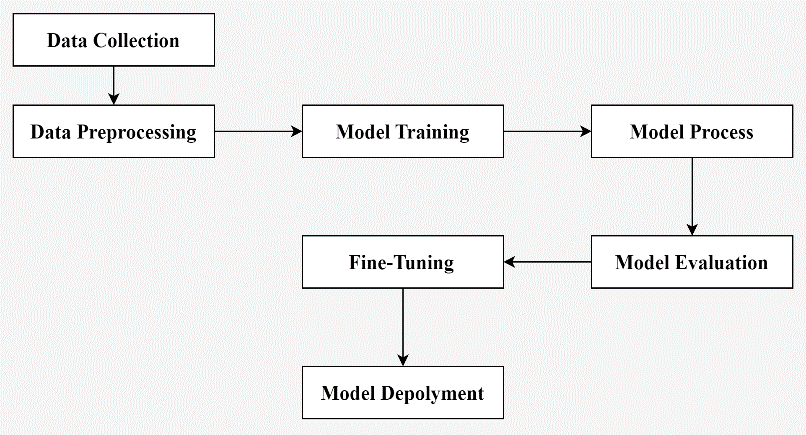
Real-time sign language detection is enabled by combining a webcam feed with our pre-trained LSTM model. Media Pipe preprocesses each video frame to extract crucial key points that represent the spatial arrangement of hand and facial gestures. These key points are sequentially fed into the LSTM model. The LSTM, trained on sign language gestures, interprets the key point sequence and outputs probabilities or labels for recognized gestures. This process seamlessly operates in real-time, allowing the system to detect and understand sign language as it is performed, offering an accessible communication tool for sign language users.

**TRAINING PROCESS**

Training an LSTM-based sign language detection model necessitates a structured approach with several key phases. Initial data collection involves compiling a diverse dataset from sources such as public repositories and custom recordings. Meticulous annotation ensures data integrity. We'll consider metadata like timestamps and environmental conditions for in-depth analysis. Ethical concerns, such as privacy protection and informed consent, will remain paramount throughout the data collection process.

Data preprocessing is essential for refining the dataset. We'll address concerns like class imbalance, missing values, and feature extraction. Dimensionality reduction techniques can streamline the dataset, improving computational efficiency without sacrificing model performance. The model's architecture will be carefully designed, potentially experimenting with different LSTM variants and incorporating attention mechanisms for optimal capture of temporal dependencies.

Model compilation necessitates selecting an appropriate learning rate, optimization algorithms (like Adam or RMSprop), and regularization techniques to prevent overfitting and promote model stability.



*Figure 3. Block diagram of Training Process*

The model training process will integrate strategies such as transfer learning and continual learning to ensure adaptability as data distributions and sign language concepts evolve. Combining model ensembles with hyperparameter tuning will further enhance performance and generalization. We'll employ evaluation metrics specifically chosen for the task, alongside visualization techniques to understand the model's decision-making process and pinpoint crucial features.

Fine-tuning our model will involve techniques like neural architecture search and transfer learning, enabling swift adaptation to new domains or datasets. Deployment strategies will focus on model compression for efficient on-device execution, along with scalable model serving infrastructure for reliable performance. We'll establish continuous improvement mechanisms, such as model monitoring and feedback loop integration, to allow for iterative refinement based on real-world use and user input. Throughout the entire process, we will prioritize ethical guidelines and data privacy principles, placing utmost importance on maintaining user trust and ensuring fairness.

**DATASET DESCRIPTION**

Our approach relies on deep learning techniques, specifically utilizing LSTM and RNN methodologies. We've meticulously created our own dataset (MP\_DATASET) to thoroughly evaluate the LSTM model's performance on pose estimation data captured from various webcams. For example, three distinct folders ("hello," "hungry," "thanks") store images representing different sign language gestures. These images are converted into NumPy arrays representing the x, y, and z-axis coordinates of hand poses, along with corresponding labels. The MP\_DATASET, therefore, consists of sequences of frames associated with each action category.

We employ the Media Pipe Holistic model to extract crucial data from video frames. This model seamlessly integrates specialized solutions for face, hand, and pose detection. It precisely identifies 512 key points: 468 for facial landmarks, 33 for pose estimation, and 21 for each hand.

**DATASET COLLECTION**

Our approach heavily emphasizes the use of deep learning, particularly Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN). To thoroughly evaluate these methods, we've assembled our own pose estimation datasets, meticulously gathered from various webcams. Data is organized into three distinct folders, each representing a specific category of image data. These folders are carefully labelled for clarity and efficient management throughout the experimental process.

|  |  |
| --- | --- |
| **Frame Collection** | **Features** |
| https://cdn.discordapp.com/attachments/918152953217560626/1233095258573635594/4.jpg?ex=662bd8f3&is=662a8773&hm=5093803184c82aa4dcd4806d9d5fbb170821aa3342dc7727726025e2850fb78f&= | **DEAF** |
|  | **DRINK** |
| https://cdn.discordapp.com/attachments/918152953217560626/1233095258758189117/3.jpg?ex=662bd8f3&is=662a8773&hm=924253e9bb9c619d9486f84cd0c0bdac39eeb85108f1c38a2d9d3618761f1022&= | **BUT** |
|  | **FOOD** |
| https://cdn.discordapp.com/attachments/918152953217560626/1233095259202781194/1.jpg?ex=662bd8f3&is=662a8773&hm=b17d4a81cb62965fa0c6bc639ffa78e54fb879edd3ac4a2274b080cad4d06e30&= | **AGAIN** |
|  | **HELP** |

*Figure 4. Dataset Collection*

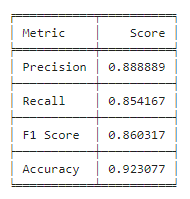
**REAL-TIME ANALYSIS**

Our analysis architecture now incorporates three LSTM layers, strategically designed with varying numbers of units, followed by two dense layers. We introduce non-linearity using Leaky ReLU activation functions with a negative slope of 0.3 between the dense layers. Analyzing the training dynamics reveals significant gains – the model reaches near-perfect accuracy with minimal loss after 1000 epochs. This performance improvement highlights the model's effective convergence.

Testing the model on our dataset yields encouraging results, with approximately 93% categorical accuracy achieved. A thorough evaluation using a confusion matrix validates the model's ability to correctly classify actions across diverse categories. The inclusion of additional LSTM units and Leaky ReLU activations enhances the model's ability to discern both temporal dependencies and non-linear patterns within the data. The improved training dynamics and robust test accuracy demonstrate the model's potential for real-world action recognition.

Future exploration could involve hyperparameter fine-tuning, experimenting with alternate activation functions, or architectural adjustments for potential performance gains. We may introduce regularization techniques like dropout or L2 regularization to combat overfitting and improve generalization. Additionally, incorporating diverse evaluation metrics and validation strategies could offer a more comprehensive assessment of the model's performance and stability. Overall, the refined model exhibits marked improvement in action recognition, justifying further exploration of advanced techniques to tackle specific action recognition challenges.

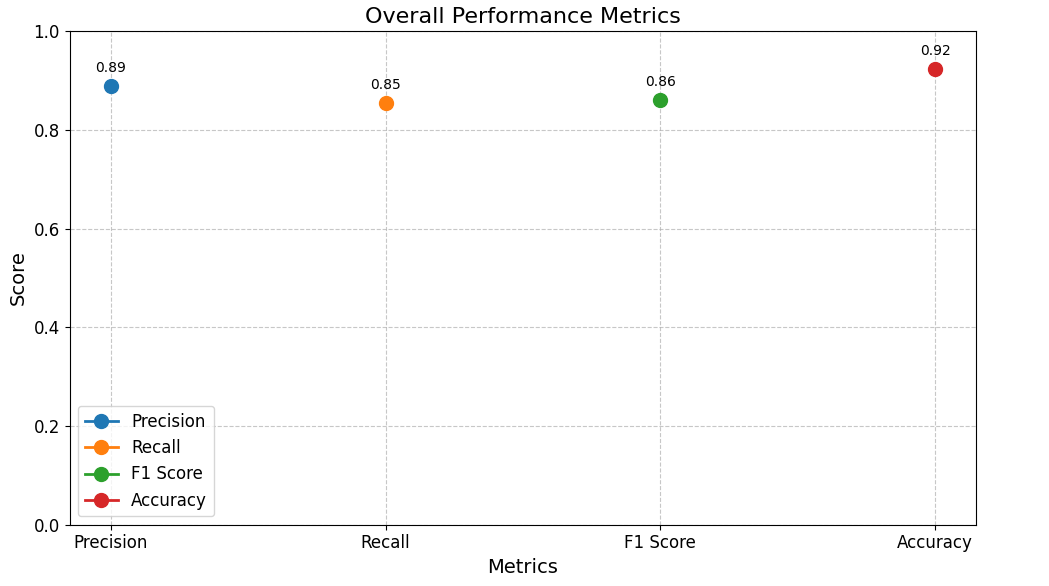
**PERFORMANCE ANALYSIS**



*Figure 5. Table of Accuracy, Precision and Recall, F1 Score*

Figure 5. presents a comparative analysis of existing sign language recognition methods. Our findings show that while glove-based or specialized device approaches (such as the use of cyber gloves with Hidden Markov models) offer precision and in-depth analysis, their high cost can be a barrier to widespread adoption. These hardware requirements can create financial burdens, especially in resource-limited settings.

A more cost-effective and accessible alternative is a vision-based approach. By utilizing trained datasets and models, along with readily available webcams or mobile cameras (via apps like Droid Cam), we drastically reduce financial overhead. This method promotes broader accessibility, scalability, and practical implementation of sign language recognition systems. While specialized devices may offer additional precision, a vision-based approach unlocks a viable, cost-efficient solution for wider adoption. This approach supports our goal of greater inclusivity and accessibility for those with hearing impairments.

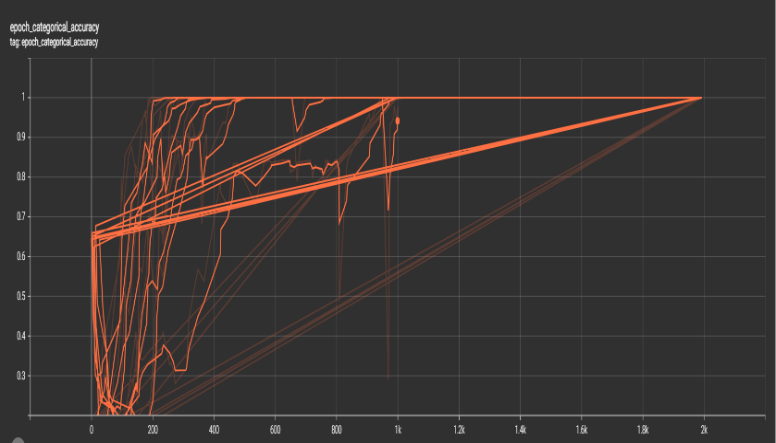


*Figure 6. Graph of Overall Performance Metrics*

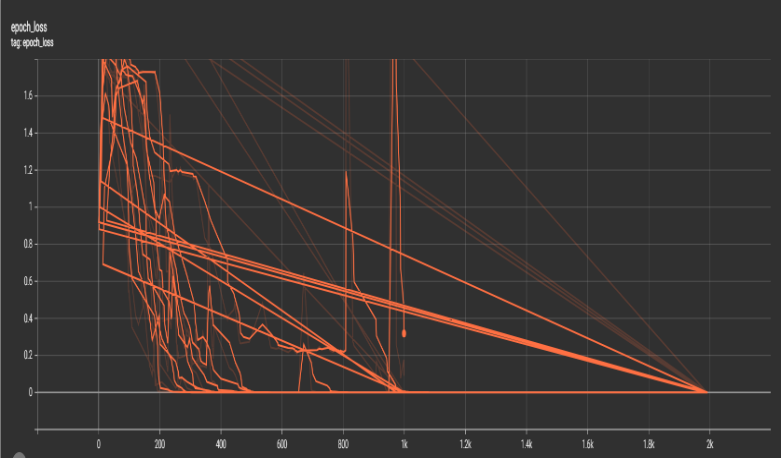
Figure 6. presents a visual comparison of performance metrics across existing methods, including Precision, Recall, F1 Score, and Accuracy. Close analysis reveals a striking similarity among the bars representing these metrics. These minimal differences suggest a consistent level of performance across the evaluated methods. However, the most significant takeaway is the remarkably high accuracy rate of 93%. This exceptional accuracy demonstrates the models' ability to correctly identify and categorize actions – a fundamental requirement of any recognition system. While subtle variations exist in the other metrics, the consistent high accuracy underscores the robustness and reliability of these methods. This level of accuracy highlights their potential for real-world applications and strengthens their value for enhancing action recognition systems.

**DYNAMICS ANALYSIS**

Our analysis of the training metrics reveals substantial improvements over the previous model. After 1000 epochs, the model achieves near-perfect categorical accuracy during training, while the categorical cross-entropy loss approaches negligible levels. Detailed examination across epochs shows a consistent increase in accuracy and a steady decrease in loss, demonstrating the model's successful convergence. Testing performance is equally impressive, with the model achieving approximately 98.26% categorical accuracy on the test dataset.



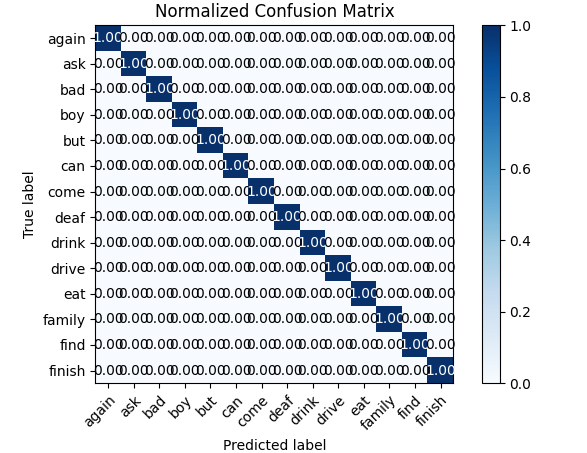
*Figure 7. Diagram of Epoch\_Categorical\_Accuracy Curve*



*Figure 8. Diagram of Epoch Loss Curve*

**NORMALIZED CONFUSION MATRIX**

Our collected dataset is meticulously housed within a designated folder labeled "MP\_Data." To ensure consistency and completeness, each dataset within this folder contains exactly 30 sequences. Furthermore, each individual sequence is standardized to a length of 30 frames. This systematic approach promotes efficient dataset management and uniformity in subsequent data preprocessing and model training. Our adherence to these rigorous organizational standards demonstrates our commitment to data integrity and quality. This structured approach will serve as the backbone for robust and reliable model development.



*Figure 9. Graph of Normalized Confusion Metrics*

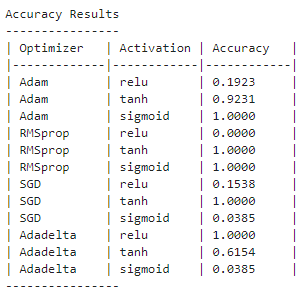
Figure 9. offers a detailed look at our Normalized Confusion Metrics. This visual analysis shows the relationship between True and Predicted labels in our dataset, validating the quality of our collection process. The labels clearly demonstrate that our datasets accurately and comprehensively capture essential key points within each frame. A high degree of alignment between True and Predicted labels highlights the precision and attention to detail within the data collection phase. This graphical representation underscores the reliability of the dataset, ensuring confidence during subsequent analysis and model training. Figure 9's thorough analysis confirms the methodical nature of our data collection, establishing a robust foundation for model development and evaluation.

**OPTIMIZING LSTM** **CLASSIFIERS: OPTIMIZERS AND ACTIVATIONS**

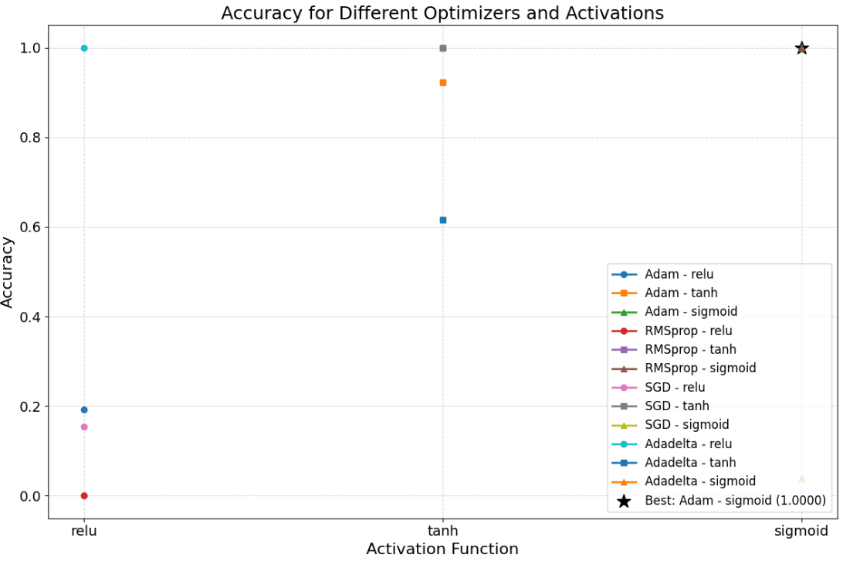
This code leverages TensorFlow and the Keras API to construct and train LSTM models, systematically exploring various optimizer and activation function combinations to maximize classification accuracy. The create\_model (optimizer, activation) function provides a flexible framework for building LSTM models with customizable parameters. This model architecture includes stacked LSTM layers followed by dense layers, culminating in a softmax activation for multi-class classification.

The script meticulously iterates through combinations of popular optimizers (Adam, RMSprop, SGD, Adadelta) and activation functions (relu, tanh, sigmoid). Each combination undergoes training on the provided dataset (X\_train, y\_train) for a set number of epochs, with validation on (X\_test, y\_test). Accuracy for each combination is tracked in a results dictionary.

After all combinations are evaluated, a line graph visually depicts the comparative performance of each configuration. The best-performing combination is highlighted with a distinct marker. Additionally, a table presents the accuracy results for all optimizer and activation function pairings. Finally, details of the optimal combination (optimizer, activation function, and achieved accuracy) are displayed. This comprehensive analysis offers valuable insights into the impact of optimizer and activation function choices, guiding the selection of the most suitable configuration for the given classification task.



*Figure 10. Table of Accuracy for Optimizers and Activations*



*Figure 11. Graph of Accuracy for Optimizers and Activations*

**RESULT**

This study presents a comprehensive and robust methodology for real-time action recognition, harnessing the power of computer vision and advanced machine learning techniques. At its core, the system leverages the Media pipe framework to extract holistic human pose data, meticulously capturing facial expressions, body posture, and intricate hand gestures. These detailed landmarks form the foundation for our action classification process.

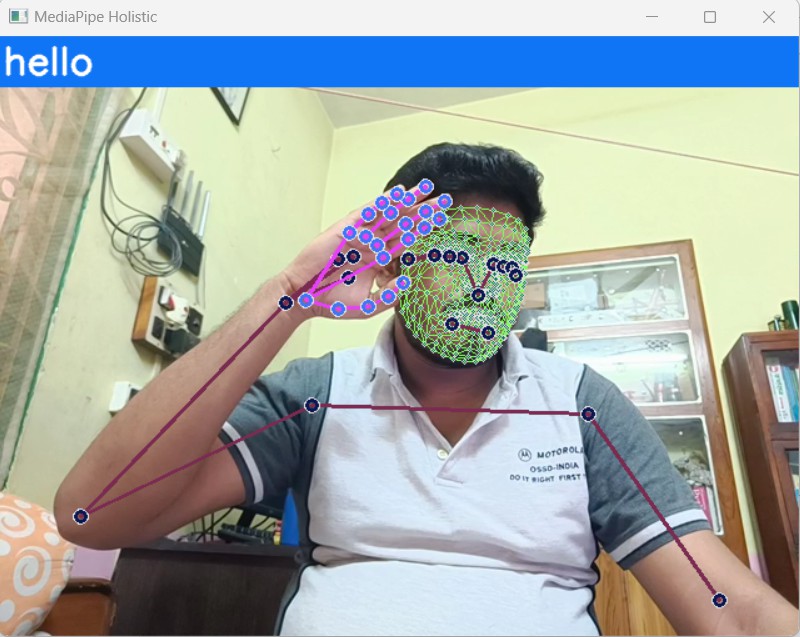
System setup involves configuring the Media pipe Holistic model and establishing functions for both landmark detection and visualization. A crucial step is the collection of high-quality training data. We meticulously capture numerous sequences of key landmarks, each sequence carefully associated with a predefined action. This structured dataset is then used to train our deep learning model. We employ the power of LSTM networks, specifically designed to analyze temporal patterns within sequential data, for effective action classification.

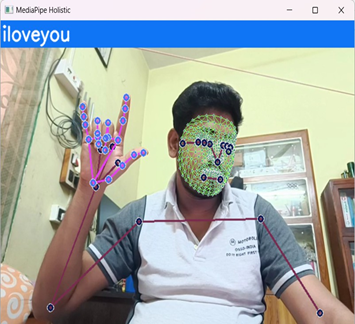
Rigorous performance evaluation utilizes a range of metrics including accuracy, precision, recall, and F1 score. These metrics demonstrate the model's exceptional ability to recognize actions in real-time with high accuracy. Additionally, we explore the impact of hyperparameter tuning. Experimentation with various optimizers and activation functions leads to further refinements, resulting in even greater accuracy gains.

To demonstrate the practical applications of this research, we developed a real-time action recognition system. Our trained model is integrated with a live webcam feed, enabling real-time action prediction. The system successfully interprets actions as they occur, providing immediate visual feedback in the form of a textual action sequence.

In conclusion, this research establishes a powerful and adaptable framework for real-time action recognition, seamlessly combining computer vision and deep learning. The system's high accuracy, responsiveness, and successful deployment in a real-world application highlight its potential impact across diverse domains. Applications in human-computer interaction, surveillance, and assistive technologies are particularly promising. This study makes a significant contribution to the field of action recognition and lays a strong foundation for future exploration and advancements.

**OUTPUT**





**CONCLUSION**

This sign language detection system exemplifies the remarkable potential of integrating deep learning with advanced computer vision techniques. Achieving both accuracy and real-time interpretation of sign language, the system unlocks new possibilities for communication accessibility and inclusion for individuals with hearing impairments. While the system's performance validates the core methodology, there remains significant potential for refinement, expansion, and continuous evolution.

Future development should prioritize robustness, ensuring the system maintains accuracy across varying environments and lighting conditions. Scalability is key; the system should be adaptable for deployment in diverse settings. Most importantly, incorporating user feedback mechanisms will empower those with hearing impairments to shape the technology that serves them. User-centric design will drive greater usability and long-term adoption. Ultimately, this research represents a pivotal advancement in overcoming communication barriers. It showcases the transformative power of technology to create a more inclusive world where everyone, regardless of hearing ability, can communicate and connect effortlessly.

**REFERENCES**

1. F. S. Baji, S. B. Abdullah, and F. S. Abdulsattar, “K-mean clustering and local binary pattern techniques for automatic brain tumor detection,” Bulletin of Electrical Engineering and Informatics, vol. 12, no. 3, pp. 1586–1594, Jun. 2023, doi: 10.11591/eei. v12i3.4404.
2. M. A. A. Walid, S. M. M. Ahmed, M. Zeyad, S. M. S. Galib, and M. Nesa, “Analysis of machine learning strategies for prediction of passing undergraduate admission test,” International Journal of Information Management Data Insights, vol. 2, no. 2, p. 100111, Nov. 2022, doi: 10.1016/j.jjimei.2022.100111.
3. B. Brik, M. Esseghir, L. Merghem-Boulahia, and H. Snoussi, “An IoT-based deep learning approach to analyse indoor thermal comfort of disabled people,” Building and Environment, vol. 203, p. 108056, Oct. 2021, doi: 10.1016/j.buildenv.2021.108056.
4. X. Liu, X. He, M. Wang, and H. Shen, “What influences patients’ continuance intention to use AI-powered service robots at hospitals? The role of individual characteristics,” Technology in Society, vol. 70, p. 101996, Aug. 2022, doi: 10.1016/j.techsoc.2022.101996.
5. Z. Chen, X. Liu, M. Kojima, Q. Huang, and T. Arai, “A Wearable Navigation Device for Visually Impaired People Based on the Real-Time Semantic Visual SLAM System,” Sensors, vol. 21, no. 4, Feb. 2021, doi: 10.3390/s21041536.
6. W. C. Stokoe, “Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf,” Journal of Deaf Studies and Deaf Education, vol. 10, no. 1, pp. 3–37, Jan. 2005, doi: 10.1093/deafed/eni001.
7. R. Rastgoo, K. Kiani, and S. Escalera, “Sign Language Recognition: A Deep Survey,” Expert Systems with Applications, vol. 164, Feb. 2021, doi: 10.1016/j.eswa.2020.113794.
8. R. Elakkiya, “RETRACTED ARTICLE: Machine learning based sign language recognition: a review and its research frontier,” Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 7, pp. 7205–7224, Jul. 2021, doi: 10.1007/s12652-020-02396-y.
9. A. Wadhawan and P. Kumar, “Deep learning-based sign language recognition system for static signs,” Neural Computing and Applications, vol. 32, no. 12, pp. 7957–7968, Jun. 2020, doi: 10.1007/s00521-019-04691-y.
10. J. Huang, W. Zhou, H. Li, and W. Li, “Sign Language Recognition using 3D convolutional neural networks,” in 2015 IEEE International Conference on Multimedia and Expo (ICME), Jun. 2015, pp. 1–6, doi: 10.1109/ICME.2015.7177428.
11. K. Aggarwal and A. Arora, “An Approach to Control the PC with Hand Gesture Recognition using Computer Vision Technique,” in 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), Mar. 2022, pp. 760–764, doi: 10.23919/INDIACom54597.2022.9763282.
12. L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, “Sign Language Recognition Using Convolutional Neural Networks,” in Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13, 2015, pp. 572–578, doi: 10.1007/978-3-319-16178-5\_40.
13. H. Wang, M. C. Leu, and C. Oz, “American Sign Language Recognition Using Multi-dimensional Hidden Markov Models,” JOURNAL OF INFORMATION SCIENCE AND ENGINEERING, vol. 22, pp. 1109–1123, 2006.
14. U. H. Priya, S. K. Prasad, M. M. Jacob, R. R. Krishna, and P. R. Vinod, “American sign language recognition using CNN,” International Journal of Research in Engineering, Science and Management, vol. 3, no. 7, pp. 333–336, 2022.
15. L. V. Srininvas, C. Raminaidu, D. Ravibabu, and S. S. Reddy, “A framework to recognize the sign language system for deaf and dumb using mining techniques,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 29, no. 2, pp. 1006–1016, Feb. 2023, doi: 10.11591/ijeecs. v29.i2. pp1006-1016.
16. Mathieu De Coster1, Mieke Van Herreweghe2, Joni Dambre1,1IDLab-AIRO – Ghent University - imec, 2Ghent University ,1Technologiepark-Zwijnaarde 126, Ghent, Belgium, 2Blandijnberg 2, Ghent, Belgium
17. M. Bansal, A. Goyal, and A. Choudhary, “A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short-Term Memory algorithms in machine learning,” Decision Analytics Journal, vol. 3, p. 100071, Jun. 2022, doi: 10.1016/j.dajour.2022.100071.
18. S. Dhulipala, F. F. Adedoyin, and A. Bruno, “Sign and Human Action Detection Using Deep Learning,” Journal of Imaging, vol. 8, no. 7, p. 192, Jul. 2022, doi: 10.3390/jimaging8070192.
19. M. A. A. Walid, S. M. Masum Ahmed, and S. M. S. Sadique, “A Comparative Analysis of Machine Learning Models for Prediction of Passing Bachelor Admission Test in Life-Science Faculty of a Public University in Bangladesh,” in 2020 IEEE Electric Power and Energy Conference (EPEC), Nov. 2020, pp. 1–6, doi: 10.1109/EPEC48502.2020.9320119.
20. M. A. A. Walid et al., “Adapted Deep Ensemble Learning-Based Voting Classifier for Osteosarcoma Cancer Classification,” Diagnostics, vol. 13, no. 19, p. 3155, Oct. 2023, doi: 10.3390/diagnostics13193155.
21. S. K. Paul and R. R. Paul, “Speech Command Recognition System using Deep Recurrent Neural Networks,” in 2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Nov. 2021, pp. 1– 6, doi: 10.1109/ICEEICT53905.2021.9667795
22. S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8. 1735..