IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

American Sign Language Recognition Based On Machine Learning And Deep Learning

[1]Ms.Kavita Sahebrao Shinde, [2]Dr.Brijendra Gupta

[1]PG Student, [2]Associate Professor

[1]Department of Computer Engineering,

[1]Siddhant College Of Engineering, Sudumbare, Pune, India.

Abstract: American Sign Language (ASL) recognition has achieved significant attention due to its potential to bridge the communication gap between the hearing and deaf communities. Recent advancements in machine learning (ML) and deep learning (DL) have revolutionized ASL recognition systems, enabling more accurate and efficient interpretation of hand gestures, facial expressions, and body movements. This study explores state-of-the-art approaches to ASL recognition using ML and DL techniques, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer-based architectures. The paper discusses key challenges such as variability in signing styles, complex gestures, and dynamic movement, while addressing solutions such as data augmentation, transfer learning, and multi-modal integration of vision and sensor-based data. Furthermore, the incorporation of attention mechanisms and real-time processing capabilities is examined to improve system adaptability and usability. Experimental results demonstrate that these methodologies achieve high recognition accuracy, with potential applications in real-time translation devices and educational tools. This research underscores the transformative role of ML and DL in enhancing accessibility and inclusivity for the deaf and hard-of-hearing community.

Index Terms - Sign Language Recognition; Manifold learning; Machine learning; Deep learning; CNN; Dimension reduction.

I. INTRODUCTION

There are so many disabilities have a strong demand to find a proper way to conveniently communicate with others. According to the WHO, 300 million people are deaf, 285 million are blind and 1 million are mute. Nowadays, Artificial intelligence is widely used in various industry, especially in Image Recognition. Therefore, this study intends to solve this problem based on machine learning and deep learning. There are few applications for mute that uses sign language to transfer sign language to words or audio to communicate with others. American Sign Language (ASL) is a rich and complex visual language used predominantly by the deaf and hard-of-hearing communities. It relies on hand gestures, facial expressions, and body postures to convey meaning. With the growing need for accessible communication technologies, automatic ASL recognition has become an essential research domain. Machine learning (ML) and deep learning (DL) techniques are at the forefront of this innovation, offering robust solutions for real-time recognition and translation.

Sign language recognition (SLR) in AI typically relies on a combination of computer vision techniques and machine learning algorithms. The specific algorithm depends on the task (e.g., gesture recognition, word recognition, or sentence recognition) and the input data type (video or images). Here are the key components and algorithms commonly used:

1) Feature Extraction : Convolutional Neural Networks (CNNs)**:** Used to extract spatial features from images or video frames. Algorithms like OpenPose or MediaPipe detect human body landmarks (hands, face, body) to focus on the gestures without processing the entire frame. Tracks motion patterns between frames to identify dynamic gestures.

2) Temporal Modeling (Sequence Processing)Recurrent Neural Networks (RNNs)**:** Models such as Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRUs) process sequences of frames for time-series data. Temporal Convolutional Networks (TCNs) Used as an alternative to RNNs for capturing temporal dependencies. Modern architectures like Vision Transformers (ViTs) and their extensions can handle both spatial and temporal information effectively.

3) End-to-End Models: 3D CNNs**:** Extract both spatial and temporal features directly from video data (e.g., C3D, I3D architectures). Spatio-Temporal Graph Neural Networks (ST-GNNs)**:** Analyze relationships between detected keypoints over time, particularly for hand and body movements.

4) Classifiers: Support Vector Machines (SVMs)**:** Used in earlier systems for static gesture recognition. Preprocessing and Data Augmentation: Segmentation: Identifies and isolates hands or relevant regions from the background.Normalization**:** Aligns gesture videos by scale, rotation, and position for consistency.

5) Datasets and Transfer Learning: Transfer learning from models pre-trained on large gesture datasets helps accelerate training and improve performance on smaller sign language datasets.

In ASLR common algorithms such as PCA, Random Forest Classification (RFC), Deep Neural Network (DNN), CNN, Data Augmentation, Manifold Learning, KNN, Gaussian Naïve Bayes (GNB), SVM, and Stochastic Gradient Descent (SGD) were used. And observe their accuracy, error, loss, and other key values, organize an objective experimental report.

Why ASL Recognition?

1) Bridging Communication Gaps: ASL recognition systems facilitate communication between hearing and non-hearing individuals by translating sign language into text or speech.
2) Increased Accessibility: These systems support inclusive technology, enabling wider access to services for the deaf community.
3) Educational Tools: Automated recognition aids in teaching ASL to learners by providing instant feedback and learning resources.

II. RESEARCH METHODOLOGY

A) Dataset Description and Data preprocessing

The Sign Language MNIST dataset **is** a dataset specifically designed for training machine learning models to recognize American Sign Language (ASL) letters. It is modeled after the structure of the MNIST dataset for handwritten digits and is widely used for research and educational purposes in computer vision and machine learning. The dataset includes grayscale images of size 28×2828 \times 2828×28 pixels. Each image corresponds to one of 24 hand gestures representing ASL letters. (Letters 'J' and 'Z' are excluded because they involve motion.) The labels are encoded as integers ranging from 0 to 23, each corresponding to a letter of the ASL alphabet. Approxiamately dataset contains Training set: ~27,455 images & Test set: ~7,172 images.

Figure 1. Sign language MNIST dataset.

Typical Workflow:

1. Data Collection: Collect gesture data using cameras, sensors, or gloves.
2. Preprocessing: Normalize, segment, or augment the data.
3. Feature Extraction: Use PCA, manifold learning, or CNNs for feature extraction.
4. Model Training:
   - For smaller datasets, try KNN, SVM, or GNB.
   - For larger datasets, use RFC, DNN, or CNNs.
5. Evaluation: Measure accuracy, precision, recall, and F1 score on a test set.
6. Fine-Tuning: Adjust hyperparameters and retrain for better performance. Each method has its strengths and weaknesses, and their effectiveness depends on the dataset size, type, and complexity of gestures

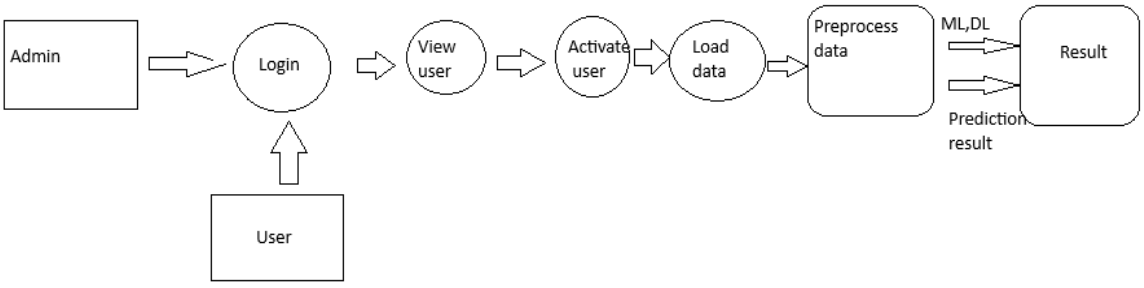

Figure 2. Overview of Proposed System

To greatly save the time of the operations of machine learning and deep learning algorithm, Principal Component Analysis (PCA) was used in this study to reduce the dimension of data. PCA get used to map high-dimensional data sets to low-dimensional space while preserving as many variables as possible. PCA get used for reducing data linearly and manifold learning get used for reducing data non-linearly. The aim of Manifold learning is to make linear frameworks like PCA to be sensitive to non-linear structure in data. In this data is dimensionality-reduced and visualized by three methods of manifold learning: MDS, T-SNE and ISOMAP. After comparing these three methods it can be observed from the graphs that the effect of using ISOMAP is the best.

Implementing various methods for American Sign Language (ASL) recognition often involves applying a combination of feature extraction, dimensionality reduction, machine learning, and deep learning techniques. Here's an explanation of how each of these techniques might be used in this context.

B) Machine learning methods:

1) PCA (Principal Component Analysis): Principal Component Analysis (PCA) is a popular technique in machine learning and data science used for dimensionality reduction**,** feature extraction, and data visualization**.** It transforms a dataset with possibly correlated features into a new set of uncorrelated features, called principal components, while retaining as much variability (information) as possible. The purpose of PCA is Dimensionality reduction. PCA is used to reduce the dimensionality of high-dimensional ASL data (e.g., hand shapes, positions, and movements captured from sensors or images). It helps in extracting the most important features while retaining variability in the data. This can enhance computational efficiency and improve classifier performance.

2) Random Forest Classification (RFC): Random Forest Classification is a machine learning algorithm based on the ensemble learning technique. It combines the predictions of multiple decision trees to improve accuracy, reduce over fitting , and enhance generalization. The purpose of Random Forest Classification is Supervised classification.
After feature extraction (e.g., PCA or other techniques), RFC can classify the extracted features into specific ASL gestures. The ensemble nature of RFC, combining multiple decision trees, provides robustness to noise and over fitting.

3) Deep Neural Network (DNN): A deep neural network (DNN) is a type of artificial neural network characterized by its deep structure, meaning it consists of multiple layers of interconnected nodes (neurons). These layers allow the network to model complex, hierarchical relationships in data. DNNs are a cornerstone of **deep learning**, a subset of machine learning focused on models with many layers. The purpose of DNN is Learning complex, high-dimensional mappings.
A DNN can directly process raw ASL data or extracted features to model the non-linear relationships between input features and output gesture labels. It is particularly effective when there's a large dataset to train on.

4) Convolutional Neural Network (CNN): Convolutional architectures in deep learning, commonly referred to as Convolutional Neural Networks (CNNs)**,** are specialized neural networks designed to process structured data like images, audio, and video. They are particularly effective for tasks such as image classification, object detection, segmentation, and more. The convolution operation is the core of CNNs. It involves sliding a small filter (or kernel) over an input (such as an image) and computing do products between the filter's weights and the local regions of the input .This operation allows CNNs to detect local patterns (e.g. ,edges, textures) in the data. For image or video-based ASL recognition, CNNs are employed to automatically learn spatial features such as hand shapes and gestures. CNN layers handle feature extraction, and fully connected layers perform classification.

5) Data Augmentation: The purpose of Data augmentation is to increase dataset size and diversity. In ASL recognition, data augmentation can involve applying transformations like flipping, rotation, scaling, adding noise, or changing lighting conditions. This helps make models robust to variations in input data.

6) Manifold Learning: The purpose of this method is **Non-linear dimensionality reduction.** Methods like t-SNE or Isomap are used to uncover lower-dimensional structures in high-dimensional ASL gesture data. These representations can enhance interpretability and classifier performance. The aim of Manifold learning is to make linear frameworks like PCA to be sensitive to non-linear structure in data. In this data is dimensionality-reduced and visualized by three methods of manifold learning: MDS, T-SNE and ISOMAP. After comparing these three methods it can be observed from the graphs that the effect of using ISOMAP is the best.

7**)** K-Nearest Neighbors (KNN): K-Nearest Neighbors (KNN) is a popular and simple machine learning algorithm used for both classification and regression tasks. Here’s an overview of its functionality, advantages, disadvantages, and some example applications in real-world machine learning projects. KNN does not involve a learning phase in the traditional sense. It simply stores the training data points. For classification, given a query point, it calculates the distances to all training points and identifies the k-nearest neighbors.

8) Gaussian Naïve Bayes (GNB): Gaussian Naive Bayes (GNB) is a probabilistic machine learning algorithm that is based on the Bayes theorem and is particularly suited for classification tasks. It assumes that the features follow a Gaussian (or normal) distribution. This algorithm is widely used because of its simplicity, efficiency, and effectiveness for certain types of problems.

Bayes Theorem**:**

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

Where:

P(C|X): Posterior probability of class C given data X.

P(X|C): Likelihood of data X given class C.

P(C): Prior probability of class C.

P(X): Evidence, constant for all classes.

9) Support Vector Machine (SVM): Support Vector Machine (SVM) is a supervised machine learning algorithm primarily used for classification and regression tasks. It works by finding the hyperplane that best separates data points belonging to different classes in a dataset. Here's a breakdown of its key features**.** A decision boundary that separates different classes in the feature space. In 2D, it's a line; in 3D, it's a plane; and in higher dimensions, it's generalized as a hyperplane. Data points closest to the hyperplane. These are the most critical points for defining the hyperplane and influencing the SVM model. The distance between the hyperplane and the closest data points (support vectors). SVM aims to maximize this margin for better generalization.

10) Stochastic Gradient Descent (SGD): Stochastic Gradient Descent (SGD) is an optimization algorithm widely used in machine learning and deep learning to train models. It is particularly effective for large-scale data and high-dimensional parameter spaces, such as those in neural networks. In standard Gradient Descent, the gradient is computed using the entire dataset, which can be computationally expensive for large datasets. SGD, instead, updates the parameters using the gradient computed from a **single** randomly chosen data point (or a mini-batch) at each iteration. This makes it faster and more memory-efficient.

III. RESULT

A. Performance for models

Table I-III indicates the performance of models in different conditions

Table I. The Results of Different Machine Learning Algorithms With Original Data.

Model Name	**Evaluation Metric**			
	test accuracy score	**test precision score**	**test recall score**	**test f1 score**
RFC	0.8161	0.80	0.81	0.80
KNN (K=1)	0.7817	0.8038	0.7817	0.7812
Gaussian NB	0.3898	0.4630	0.3898	0.3904
SVM	0.8419	0.8568	0.8419	0.8444
SGD	0.6602	0.7072	0.6602	0.6713

Table II. The Results of Different Machine Learning Algorithms With Data Processed By PCA.

Model	**Evaluation Metric**

Name				
	test accuracy score	test precision score	test recall score	test f1 score
RFC	0.087	0.09	0.09	0.09
KNN (K=1)	0.8209	0.8402	0.8209	0.8225
Gaussian NB	0.5889	0.6692	0.5889	0.6091
SVM	0.8515	0.8638	0.8515	0.8532
SGD	0.6429	0.6670	0.6429	0.6451

Table III. The Results Of Different Machine Learning Algorithms With Data Processed By Isomap

Model Name	Evaluation Metric			
	test accuracy score	test precision score	test recall score	test f1 score
RFC	0.1433	0.14	0.13	0.13
KNN (K=1)	0.9654	0.9659	0.9654	0.9654
Gaussian NB	0.0400	0.0414	0.0400	0.0352
SVM	0.0349	0.0406	0.0349	0.0304
SGD	0.0380	0.0424	0.0380	0.0348

By compairing above 3 tables following result and discussions are made.

A. Result and discussion with SVM

SVM performs better than other machine learning methods. The accuracy score of original data is almost the same as the accuracy score after doing PCA. But after doing PCA, the training time can be reduced by 10 minutes. So the main target of this paper is use PCA for diamention reduction and SVM as a classifier for achieving best result. This combination preferred when we have to reduce diamentions of data linearly.

B. Result and discussion with KNN

When using KNN to train and predict the data, we found that the accuracy is higher when the K is smaller. By cross validation, we chose the K with the highest accuracy. After that, we used the data which is dimensionality-reduced by PCA and ISOMAP and found out the accuracy is higher than before. And when using ISOMAP the accuracy increased by more than ten percent.This combination preffered when we have to reduce data non-linearly.

C. Result and discussion for Neural Networks

Table IV. The Results Of Different Machine Learning Algorithms With Data Processed By DNN and CNN.

Model Name	Evaluation Metric	
	Loss	Accuracy
DNN-1	6.3743	0.8292
DNN-2	3.7469	0.4169
CNN	1.0108	0.9387
CNN after Augmentation	0.0962	0.9781

From above table it is found that the performance of CNN is better than DNN. Different parameters affect the accuracy of model. After the adjustment of parameters, the accuracy improves, and the accuracy of training set is 0.9997 also in test set this value reaches 0.9387. The problem of over fitting has been improved. CNN is much stronger than DNN, it can effectively reduce the dimension of large data images to small data and retain the characteristics of the picture, similar to the principle of human vision. After Data Augmentation, the performance of model improves continually. The accuracy in test set reaches 0.9781, the loss in test set reduces either.

IV. CONCLUSION

In this work, American Sign Language Recognition was proposed, using many methods to train models to classify and recognize 24 gesture letters. Among the 26 letters, J and Z are excluded because they need finger movement. This study developed PCA for linearly reduce diamentions of input data and Manifold Learning for non-linearly reduce dimensions of input data and can fasten the speed of training. The performances of RFC, KNN, GNB, SVM and SGD are compared. Simultaneously, DNN and CNN models are trained to see their performances. Here in case of original data, SVM has the best effect, with test accuracy reaching 0.8419. After PCA dimensionality reduction, SVM has a good effect with 0.8515. However, after ISOMAP dimension reduction, the effect of KNN was greatly improved, reaching 0.9654. Here we seen that CNN has better effect than DNN with less loss of data. The performance of CNN improved after Data Augmentation, the accuracy of test set can reach 0.9781 finally. So here proved that CNN after Data Augmentation has the best performance. In this way we can make a reliable and fast communication system for disabled people with normal people.

V. REFERENCES

[1] M. Sharma, P. Ranjana, and K. Ashok. "Indian sign language recognition using neural networks and KNN classifiers." ARPN Journal of Engineering and Applied Sciences, vol. 9, 1255-1259, 2014.

[2] Lanxi Li , Da Liu, Chaplin Shen and Jing Sun “American Sign Language Recognition Based on Machine Learning and Neural Network.” 2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE).

[3] R. Rastogi, M. Shashank, and A. Sajan. "A novel approach for communication among Blind, Deaf and Dumb people." 2015 2nd International Conference on Computing for Sustainable Global Development (INDIA Com). IEEE, 2015.

[4] Y. Qiu, et al. "Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training." Biomedical Signal Processing and Control, vol. 72, 103323, 2022.

[5] L. Pigou, et al. "Sign language recognition using convolutional neural networks." European Conference on Computer Vision. Springer, Cham, 2014.

[6] P. Haque, D. Badhon, and N. Nazmun. "Two-handed bangla sign language recognition using principal component analysis (PCA) and KNN algorithm." 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE, 2019.

[7] "What is a KNN (K-Nearest Neighbors)", https://www.unite.ai/what-is-k-nearest-neighbors/
[8] J. Liu, J. Chen, and J. Cheng. "Static Gesture Recognition System for Human-Computer Interaction." Infrared and Laser Engineering, 2002.
[9] Tech person. "Sign Language MNIST." Kaggle, https://www.kaggle.com/datasets/datamunge/sign-language mnist.
[10] Keras. "Keras: Deep learning library based on Python". https://keras.io/zh/preprocessing/image.
[11] "How to Create a Decision Tree Classifier in Python Using Sklearn." Learning about Electronics, http://www.learningaboutelectronics.com/Articles/How-to-create-a- random-forest-classifier-Python-sklearn .php.
[12] H. Zhang, The optimality of Naive Bayes." Proc. FLAIRS, 2004.
[13] N. Amor, B. Salem, and E. Zied. "Naive Bayes vs decision trees in intrusion detection systems." Proceedings of the 2004 ACM symposium on Applied computing, 2004.
[14] S. Karamizadeh, et al. "Advantage and drawback of support vector machine functionality." 2014 international conference on computer, communications, and control technology (I4CT). IEEE, 2014.
[15] "Deep Neural Network: The 3 Popular Types (MLP, CNN and RNN)". https://viso.ai/deep-learning/deep-neural-network-three-popular-types/
[16] IBM. "Convolutional Neural networks", https://www.ibm.com/cloud/learn/convolutional-neural-networks
[17] X. Geng, D. Zhan, and Z. Zhou. "Supervised nonlinear dimensionality reduction for visualization and classification." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 35, pp.1098- 1107, 2005.