

# Machine Learning and Neural Network-driven Recognition of American Sign Language

<sup>1</sup>Syed Mujahed Husen, <sup>2</sup>Shaik Sohel, <sup>3</sup>Haroon Ahmed, <sup>4</sup>Mohammed Faisal Uddin

<sup>1</sup>Associate Professor, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd.

<sup>2,3,4</sup>B.E Student, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd.

[syed.mujahed2000@gmail.com](mailto:syed.mujahed2000@gmail.com), [shaiksohel280@gmail.com](mailto:shaiksohel280@gmail.com), [haroonahmed0384@gmail.com](mailto:haroonahmed0384@gmail.com),  
[mohammedfaisal7272@gmail.com](mailto:mohammedfaisal7272@gmail.com)

**Abstract:** Many people with disabilities consisting of the deaf and mute go through because they're not able to talk with human beings. It is important to find a manner to remedy this trouble. One feasible approach is language popularity (SLR) that is a form of pattern popularity. In this paper, gadget gaining knowledge of and deep studying are used to apprehend and classify ASL, and simplest 24 English letters are labelled because the letters J and Z require finger movements. First, important component evaluation (PCA) and numerous algorithms are used to lessen the rate of system studying and visualization. Second, various system learning strategies which include Random Forest Classification (RFC), K-Nearest Neighbour (KNN), Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM) and Stochastic Gradient Descent (SGD) are used to divide the structure. . Since the SVM algorithm has many hyper parameters, this have a look at makes use of a grid seek approach to locate the first-rate mixture of hyper parameters for more correct prediction. It has been located that the distinctive discount algorithms have unequal accuracy of every prediction version, and it can be concluded that the various set of rules is the pleasant dimension discount algorithm only is for KNN but now not for different prediction models, and PCA is extra green. Than KNN utilized in machine learning algorithms except KNN. Two deep studying methods including Convolution Neural Networks (CNN) and Deep Neural Networks (DNN) also are utilized in category and their accuracy is better than the algorithms cited above.

**Keywords-** Sign Language Recognition; Manifold; Machine learning; CNN; Dimension reduction.

## I. INTRODUCTION

According to the World Health Organization, 285 million humans are blind, three hundred million are deaf and 1 million are illiterate [1]. Many human beings with disabilities need to find methods to speak effortlessly with others. It is well known that sign language is broadly used in communication between deaf and mute human beings. Although there are numerous natural language learning strategies used by hearing and blind humans, there are few applications for sign language rather than sign language for voice or audio conversation with others.

Today, synthetic intelligence (AI) is broadly used in many sectors, mainly in image popularity. Therefore, this work plans to clear up this problem primarily based on machine learning and deep learning knowledge of.

In the sphere of AI, many algorithms can be used to realise the role of language recognition: convolution neural network (CNN), which is a form of synthetic neural network (ANN), can realise automatic feature extraction and classification. Some researchers have proposed the use of neural network and K-Nearest Neighbour (KNN) classification to classify languages [5]. Some researchers have used principal component analysis (PCA) to extract features to reduce the data quantity,

which converts high records into low statistics [6]. Furthermore, a part of the studies compares multi-layer perceptron (MLP), radial basis function (RBF), Mahalanobis distance and least squares support vector machine (LS-SVM) in language popularity [7]. In truth, there are numerous classification and dimensionality reduction strategies, including Random Forest and T-SNE, to cope with the dataset. However, it is clear that few researchers control to provide an explanation for the functioning of numerous algorithms inside the same experiment. Therefore, the cause of this text is to explain the performance of various algorithms in language recognition (SLR) and to provide a reference for researchers who want to apply those algorithms.

To attain this aim, unique techniques, or strategies including 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) are tested in this test. And cope with their truth, mistakes, losses and different crucial things, set the goal of learning. Additionally, it's far difficult to examine the overall performance of algorithms based totally on extraordinary

records because of specific angles, light, and many others. Images taken from each dataset, so the scope that these records can constitute is constrained. Considering those adjustments, it is important to apply a random dataset to check different algorithms to check their performance.

## II REVIEW OF LITERATURE

Previous researchers have targeted their work available gestures to help the hearing impaired using advanced generation with synthetic intelligence. Algorithms. Although loads of research has been done on SLR, there are nevertheless boundaries and enhancements wishes to be addressed to enhance the deaf community. This phase provides a brief overview of recent research on SLR using deep learning- and deep vision strategies. A evaluate of the literature on the problem indicates that there are many methods to clear up the hassle of spotting gestures in video use some special ways. In [1], the authors use Hidden Markov Models (HMM) to apprehend the faces of films together with Bayesian community classifiers and Gaussian bushes augmented naive Bayesian classifiers. Francois et al. [2] also posted a paper on human body recognition in films using 2D and three-D like strategies. The task involves the usage of

PCA to recognize silhouettes from a static digital camera after which using 3-D to model them for goal reputation. This technique has its risks there are intermediate gestures that could cause confusion in training and accordingly to decrease in prediction. Let's speak the evaluation of video segments the usage of neural networks which includes extracting visible information in the form of feature vectors. Neural networks face problems including hand detection, segmentation of Historical past and surroundings studies, dimming lights, occlusion, movement and function. The article via Nandy et al. [3] divide the dataset into segments, extract functions and divide using Euclidean distance and K-Neighbours. Ten Similar works with the aid of Kumud et al. [4] method the way to understand Indian Sign messages. Report images extracted from video files, pre-processed statistics, extract the main content material of the data, accompanied with the aid of the extraction of different functions, reputation and ultimately optimization. Pre-processing is accomplished by way of converting the video to a stage of RGB images. All snap shots there's a size. Skin shade segmentation is used to extract pores and skin areas, the usage of HSV. The photographs are converted into binary layout. Key frames are extracted by way of

calculating the gradient of these Pole. And the capabilities are extracted from key frames using oriental histogram. Classification becomes achieved by using Euclidean distance, Manhattan distance, chessboard distance and Mahalanobis distance.

**III METHOD**

A. Description of the dataset and priorities  
 The name of the dataset on this take a look at is Sign Language MNIST, which comes from Kaggle [8]. In the research procedure, two CSV documents are commonly used, one is the training data set of 27,455 sufferers and the alternative is primarily based on the checking out statistics of seven, 172 sufferers. There are 24 special languages in this file (not including J and Z which want to be moved). Each document has 784 pixels, representing a 28 × 28 pixel photo. The one-of-a-kind pixel values represent grey values and variety from zero to 255. Figure 1 indicates some instance photographs in the MNIST records register [8].



Figure 1. Sample images in the collected dataset.

After doing PCA, the number of features is reduced from 786 to 115. And we select the first four pixels to observe their dependency and distribution. As Figure 2 shows below, at positions other than diagonal, it represents the scatter plot of each pixel relative to the other three pixels. At the diagonal, it represents each pixel's distribution. As we can see, each pixel roughly confirms Normal Distribution, and each pixel is independent of the other pixels.

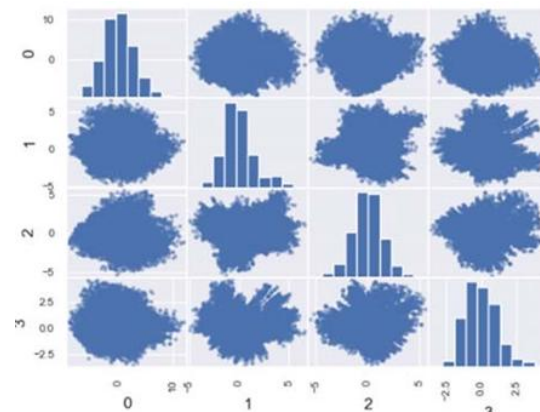


Figure 2. First four pixels distribution and independence.

The first step in statistics processing is normalization, which includes dividing every pixel with the aid of 255 to normalize the statistics (the fee of every pixel stages from 0 to at least one).

Figure three by plotting the histogram; it can be visible that all gesture patterns in the education facts seem at a similar frequency. In the test facts, a few directions appear more often, together with the frequency of label 4 and label 7 that is around 14%. The frequency of label 15 and label sixteen is about 5%. It can be visible that the uneven distribution of each label in the test information will reason a exchange inside the accuracy measurement. For instance, gestures with greater take a look at data can lessen accuracy due to the fact the opportunity of inclusive of gestures in special environments is better and different factors consisting of light Good visibility can have an effect on the picture readability. In addition, in an effort to keep time in the next processing of gadget learning and deep mastering algorithms, predominant factor evaluation (PCA) changed into used in this study to reduce the scale of the file. The layout concept of PCA is to map the high-dimensional data set right into a low-dimensional area whilst retaining as many variables as viable.

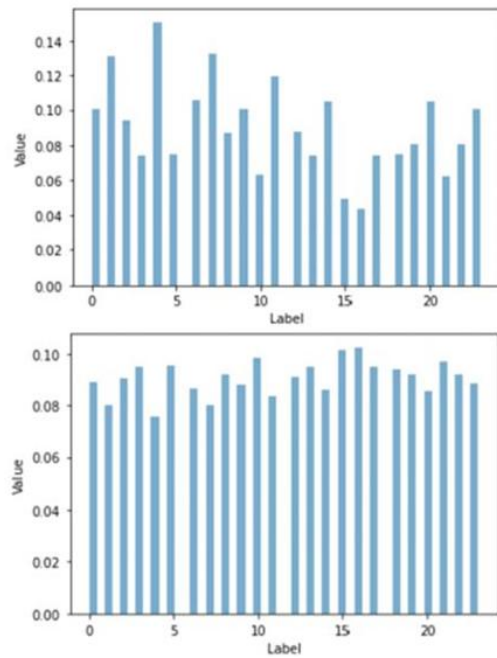


Figure 3. Distribution of gesture sample in the collected dataset.

In addition to PCA, manifold gaining knowledge of is also used to reduce facts length. Multiple mastering is a technique to decrease bad results. This can be seen as a try and visualize linear models like PCA to better apprehend non-linear patterns in the statistics. The statistics is gotten smaller and visualized via 3 more than one research: MDS, T-SNE and ISOMAP. After evaluating these 3 methods, it appears from the photos that the consequences of using ISOMAP are the fine.

#### IV RESULTS

##### 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
Gauss an 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 Name	Evaluation Metric			
	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
Gaussia nn 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
Gaussi anNB	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

Only 24 letters are educated, due to the fact J and Z require the palms to transport, but if J and Z are used in prediction, the result depends on which letter is similar.

**B. Events and discussions for the RFC**

In the RFC, we located that the accuracy is better after the discount with ISOMAP than with PCA, due to the fact ISOMAP suggests some blessings with PCA, however ISOMAP is easier to study large training of nonlinear manifolds [17] and could no longer be an excessive amount of. Linear relationship among pixels in our

training information. However, ISOMAP is not always higher than PCA. In [18], the writer examines the overall performance of linear and non-linear extraction algorithms. Feature extraction set of rules (FEA) can clear up the issues of actual records, together with noise, complexity and sparsely [18]. Studies have shown that FEA nonlinear is better in human work, however now not necessarily higher in real paintings, although nonlinear methods are proposed to triumph over the shortcomings of the method in [17]. In addition, nonlinear FEAs aren't continually better than PCAs because they may be touchy to the curse [17]. As a result, ISOMAP and PCA have extraordinary values for the curse, which makes PCA extra accurate.

**C. Results and discussion for KNN**

Using KNN to teach and expect the information, we found that the accuracy is better while K is small. By opposition, we select K with the maximum. Then, we use the facts whose length is reduced by PCA and ISOMAP and determined that the accuracy is higher than earlier than. And the usage of ISOMAP, the accuracy is improved via extra than ten percentages.

**D. Consequences and discussion for GNB**

As proven above, after completing PCA, in comparison to the original records, the

accuracy is elevated by way of 20%. Although, GNB does not paintings nicely in SLR.

To our understanding, GNB is primarily utilized in junk mail detection, but little used in sample reputation. Pixel independence and regular distribution can't be acquired in the original statistics, but can be acquired after appearing PCA. In junk mail e mail, some phrases like "charge" and "low" appear, which ends up in the belief that it's miles junk mail. In addition, the opportunity of the word performing does no longer affect other words, which means that it follows the concept of independence. On the contrary, the belief that the fee of 1 pixel does not affect the others cannot be defined well. In addition, the value of one pixel may be easy to keep away from, due to the fact PCA is a sort of straight line, every pixel can be extra or much less than others after processing PCA.

**E. Results and dialogue for SVM**

SVM plays better than different device studying strategies. The accuracy score of the authentic facts is nearly the same as the accuracy score after performing PCA. But after finishing the PCA, the education time can be decreased via 10 mines.

We use Grid Search to locate the quality aggregate of hyper parameters. We

examined several viable hyper parameter values, as proven in Table IV beneath.

TABLE IV. HYPERPARAMETERS FOR SVM

Kernel	Hyper parameter Combination		
	<i>linear</i>	<i>poly</i>	<i>rbf*</i>
C	0.1	1*	10
Degree (when kernel=poly)	2	3	4

\*means the best value

**F. Results and Discussions for Neural Networks**

To make sure the accuracy and performance of the version, the neural community does now not use the facts after reduction. According to the DNN version, for the first time, it can be seen that the accuracy of the education method is better than that of trying out, and the loss charge is inconsistent. Based on schooling history analysis, the loss fee maintains to growth on this type of neural network, which is a sign of over competitiveness.

Then reduce the variety of neurons in each layer. The hassle of over fitting still exists. As the wide variety of neurons decreases, the accuracy of schooling and testing decreases. When the quantity of neurons is 24 in each layer, the accuracy inside the schooling set is zero.758, the accuracy in the take a look at set is 0.4169, whilst the loss in the take a look at set is of 3.7469.

DNN is a simple version, its effect is not best. It is manageable that as the quantity of neurons increases, accuracy improves and loss decreases, however it is

straightforward to over fit. The overall performance of CNN is higher than that of DNN. Parameters have an effect on the accuracy of the model. After converting the parameters, the accuracy is advanced and the accuracy of the education system is zero.9997 while within the check this value reaches zero.9387. The over fitting problem has stepped forward. CNN is more effective than DNN, it is able to lessen the width of massive data to small records (does not have an effect on the consequences) and retain the photo functions, just like the unique content of human vision. After information augmentation, the version performance constantly improves. The check accuracy increased to 0.9781, the loss in the check set decreased with the aid of one.

## V CONCLUSION

In this work, language recognition is proposed, the usage of various methods to educate models to classify and apprehend 24 hand gestures. Of the 26 letters, J and Z are overlooked due to the fact they require finger motion. Those now not blanketed in these 24 businesses could be well-known as the nearest letter. This takes a look at evolved PCA and Manifold Learning to lessen the data size and speed up the evaluation. Then, the performances of RFC, KNN, GNB, SVM and SGD are as

compared. Simultaneously, DNN and CNN models are skilled to reveal their overall performance. Some experiments were performed to evaluate the proposed techniques. Manifold Learning performs better in dimensionality reduction and the dataset is decreased from 784 to 115 dimensions. Different algorithms have one of a kind overall performance, even if the algorithm is advanced after decreasing the dimensions by using extraordinary techniques. In the case of the unique records, SVM has the quality impact, with the test accuracy as much as zero.8419. After PCA measurement discount, SVM has excellent outcomes with 0.8515. However, after lowering the ISOMAP measurement, the impact of KNN is very good, accomplishing 0.9654. Through network search, this have a look at can get the best mixture of hyper parameters, which is .Kernel = rbf and C = 1. Neural networks also are powerful models, and the performance of CNN is progressed after facts augmentation, the accuracy of the check machine can reach 0.9781. By assessment, CNN after statistics augmentation has the best performance. In the future, more research could be considered to use specific evaluation strategies and use these fashions to function extra system getting to know.

## REFERENCES



1. M. Sharma, P. Ranjna, 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. P. Haque, D. Badhon, and N. Nazmun. "Two-handed bangle sign language recognition using principal component analysis (PCA) and KNN algorithm." 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE, 2019.
3. J. Liu, J. Chen, and J. Cheng. "Static Gesture Recognition System for Human-Computer Interaction." Infrared and Laser Engineering, 2002.
4. Tec person. "Sign Language MNIST." Kaggle, <https://www.kaggle.com/datasets/datamun/ge/sign-language-mnist>.
5. Keras. "Keras: Deep learning library based on Python". <https://keras.io/zh/preprocessing/image/>
6. "How to Create a Decision Tree Classifier in Python Using Sklearn." Learning about E "What is a KNN (K-Nearest Neighbours)", <https://www.unite.ai/what-is-k-nearest-neighbors/> H. Zhang, The optimality of Naive Bayes." Proc. FLAIRS, 2004.
7. 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.
8. S. Karamizadeh, et al. "Advantage and drawback of support vector machine functionality." 2014 international conference on computer, communications, and control technology (I4CT). IEEE, 2014.
9. Prasadu Peddi and Dr. Akash Saxena (2014), "EXPLORING THE IMPACT OF DATA MINING AND MACHINE LEARNING ON STUDENT PERFORMANCE", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.1, Issue 6, page no.314-318, November-2014, Available: <http://www.jetir.org/papers/JETIR1701B47.pdf>
10. Prasadu Peddi and Dr. Akash Saxena (2015), "The Adoption of a Big Data and Extensive Multi-Labeled Gradient Boosting System for Student Activity Analysis", International Journal of All Research

Education and Scientific Methods  
(IJARESM), ISSN: 2455-6211, Volume 3,  
Issue 7, pp:68-73.