IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN2347–3657

Volume13, Issue 1,2025

RECOGNITION OF AMERICAN SIGN LANGUAGE USING MACHINE LEARNING AND NEURAL NETWORKS

[1]SAYANNAGARI VAMSHI, [2]VARKOLU VENKATESH, [3]THADA AJAY, [4]MARRIPALLY VAMSHI, [5]YENUGU NITHIN REDDY, [6]Mr. M RAMA KRISHNA CHAITANYA, [7]Mr. YADAHALLI BHIMAPPA,

[12345]Student Department of DS, Narsimha Reddy Engineering College, Maisammaguda (V), Kompally, Secunderabad, Telangana-500100.

[6]Assistant Professor, Department of CSE, Narsimha Reddy Engineering College, Maisammaguda (V), Kompally, Secunderabad, Telangana-500100.

[7]Assistant Professor, Department of Mechanical Engineering, Narsimha Reddy Engineering College, Maisammaguda (V), Kompally, Secunderabad, Telangana-500100.

Abstract—

We must discover a solution to the communication barrier that prevents many persons with impairments, including the deaf and mute, from interacting with the general public. As a pattern recognition methodology, Sign Language Recognition (SLR) offers a potential solution. Only 24 English letters are recognized and categorized in this research using machine learning and deep learning approaches to detect and classify American Sign Language (ASL). This is due to the fact that letters J and Z need finger movement. To speed up machine learning training and visualization, dimension reduction using Principal Component Analysis (PCA) and manifold techniques is the first step. To categorize the pattern, many machine learning algorithms are used, including Random Forest Classification (RFC), K-Nearest Neighbor (KNN), Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM), and Stochastic Gradient Descent (SGD). This work use the Grid Search approach to determine the optimal combination of hyperparameters for the SVM algorithm, resulting in more accurate predictions. The results show that various dimensionality reduction algorithms have different impacts on the prediction models' accuracy. Specifically, we find that the manifold algorithm is the best dimensionality reduction algorithm for KNN, but not for other prediction models. When it comes to machine learning algorithms, principal component analysis (PCA) is far more practical than KNN. Among the aforementioned techniques, the two most accurate deep learning options for classification are deep neural networks (DNN) and convolutional neural networks (CNN).

Keywords; Sign Language Recognition; Manifold; Machine learning; CNN; Dimension reduction

Introduction

Worldwide, 285 million individuals lack the ability to see, 300 million are hard of hearing, and 1 million cannot speak a word, records show [1]. Finding an appropriate and easy method of communicating with others is a pressing need for many people with disabilities. The widespread use of sign language as a means of communication among the deaf and dumb is well-known. While many studies have focused on the ways in which persons with normal hearing and vision use language, very few have explored ways for those who are deaf or visually impaired to translate sign language into spoken language.

These days, many different industries rely on artificial intelligence (AI), with image recognition being one of the most prominent. This research aims to address that issue by using ML and DL to the topic at hand.

The purpose of sign language recognition may be accomplished with the help of several algorithms developed in the area of artificial intelligence: Automated feature extraction and categorization is possible using Convolutional Neural Networks (CNNs), a kind of Artificial Neural Networks (ANNs) [2-4]. When asked about methods for classifying sign language, some academics brought up neural networks and K-Nearest Neighbor (KNN) classification [5]. Data that was previously high dimensionality was transformed into low dimensionality by certain researchers using Principal Component Analysis (PCA) to extract characteristics [6]. Another area of research that analyzes several methods for sign language identification is the Mahala nobis distance, Least Square Support Vector Machine, Radial Basis Function, and Multilayer Perceptron (MLP) [7]. Data sets may be processed using dimensionality-reduced methods and additional classification algorithms like T-SNE and Random Forest. But it turns out that not many academics can explain how different algorithms performed in the same test. Thus, this paper's goal is to shed light on how various algorithms perform in sign language recognition (SLR) and serve as a resource for academics interested in using these methods.

Various standard techniques and strategies were used in this experiment to reach this goal: 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). Create an unbiased report detailing your experiments while keeping track of their precision, inaccuracy, loss, and other critical metrics. In addition, the scope that these publications can portray is restricted due to the varying angles, lighting, etc. of the photos taken in each data set, making it impossible to evaluate the performance of algorithms based on many articles. In order to evaluate the efficacy of various algorithms in light of these factors, a standard data set must be used.

Method

Datasetdescriptionandpre-processing

This research makes use of the Sign Language MNIST data collection, which is sourced from Kaggle [8]. A training data set of 27,455 instances and a test data set of 7,172 cases are the two primary CSV files used in the research procedure. Except for J and Z, which need movement, this dataset contains 24 distinct sign languages. Every piece of data consists of 784 pixels, which is equivalent to one 28×28 pixel picture. Variegated pixel values, which range from 0 to 255, stand for grayscale values. Figure 1 showcases a selection of pictures from the Sign Language MNIST dataset that was gathered [8].

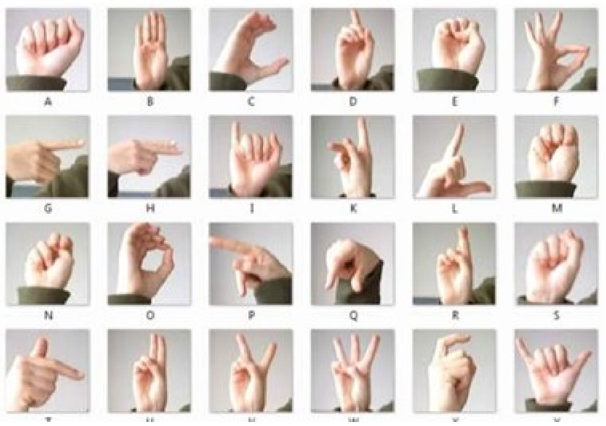

Figure1.Sampleimages inthecollecteddataset.

There are now 115 features instead of 786 after PCA. In order to see how the first four pixels are distributed and dependent on each other, we choose them. Below, in Figure 2, we can see that it depicts the scatter plot of every pixel in relation to the other three pixels at places other than the diagonal. On the diagonal, you can see the distribution of each pixel. Every pixel seems to follow a normal distribution, and they are completely separate from one another.

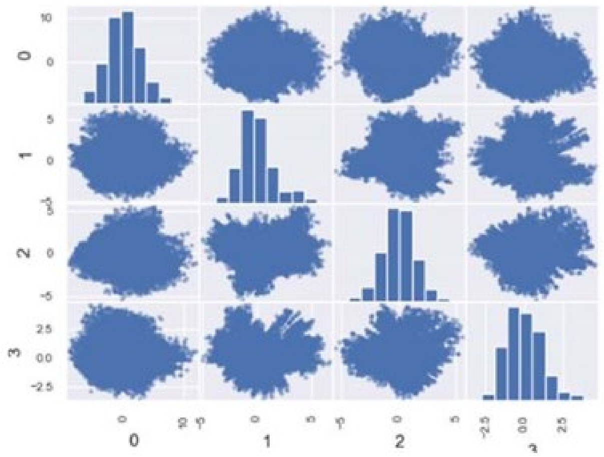

Figure2.Firstfourpixelsdistributionandindependence.

Since the values of the pixels may only take on values between 0 and 1, the first thing to do when processing the data is to normalize it by dividing it by 255. Fig. 3 The histogram shows that all of the gesture samples in the training set occur at about the same frequency. Some gestures occur more often in the test data set than others; for example, labels 4 and 7 appear roughly 14% more often. About 5% of the time, you'll see labels 15 and 16. It is clear that variations in testing accuracy might result from the test data set's unequal distribution of labels. As an example, the likelihood of adding movements recorded in diverse locations increases with more test data, which might lead to decreased accuracy. This is because external conditions, such lighting, can impact the quality of picture recognition. Additionally, this research used Principal Component Analysis (PCA) to minimize the data dimensions, which significantly sped up the following operations of the machine learning and deep learning algorithms. Reducing high-dimensional data sets to low-dimensional space while retaining as many variables as feasible is the design principle of principal component analysis (PCA).

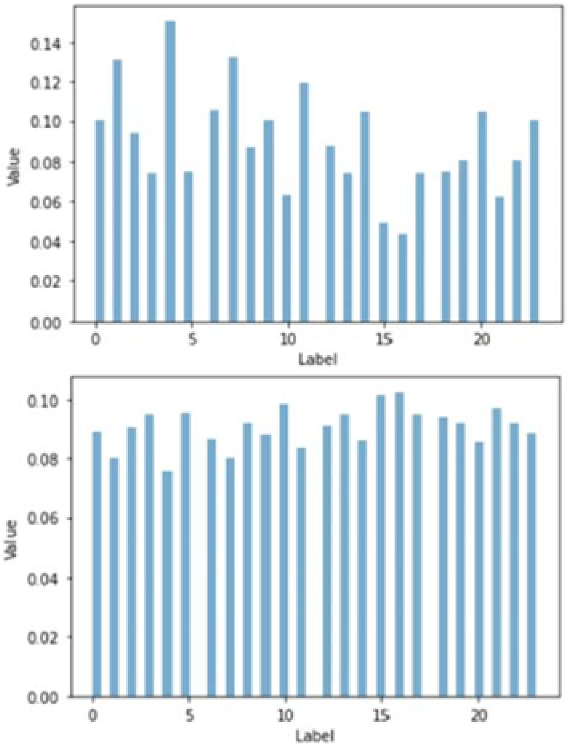

Figure3.Distributionofgesturesampleinthecollecteddataset.

Both principal component analysis (PCA) and manifold learning were used to decrease the data's dimensionality. One method for reducing dimensionality that is not linear is manifold learning. One way to look at it is as an effort to make linear frameworks, such as principal component analysis (PCA), more adaptable to data with non-linear structure. Three manifold learning methods—MDS, T-SNE, and ISOMAP—are used to decrease data dimensionality and show it. From the graphs, it is clear that ISOMAP produces the best result after comparing the three approaches.
It seems reasonable to use Data Augmentation to horizontally flip gesture graphics for left-handed persons. According to [9], ImageDataGenerator in Keras can do this. It also makes the data set larger and more variable, so the model may utilize more data while ignoring irrelevant information.throughout the training process. As a result, accuracy and resilience are both enhanced.

The first is ML models.

One, RFC. For better prediction accuracy and over-fitting management, random forest employs averaging after fitting many decision tree classifiers on separate subsamples of the dataset [10]. One kind of supervised learning is the random forest classifier, which makes predictions based on historical data. We used both training and test data sets in this investigation. After that, the model is built, trained, and finally, the prediction results are produced.

K-nearest neighbours. When it comes to regression and classification, K-Nearest Neighbors is your go-to machine learning method [11]. To determine what category a given data point belongs to, K-Nearest Neighbors looks at the labels of a subset of the data points immediately around it. In this work, KNN is used for training and prediction purposes using both the raw data and data processed with dimensionality reduction. Using cross validation and a value for k between 1 and 10, we find that, when working with the Sign Language MNIST dataset, a lower value for K yields better accuracy. Consequently, the forecast will be made using the model with the greatest accuracy. #3. GNB. According to [12], the Naïve Bayes method is a kind of supervised learning algorithm that utilizes Bayes' theorem for classification purposes, supposing, with the "Naïve" assumption, that for any given label, each pair of features functions independently. Assuming that x1–xn are characteristics of the set y and n, the following connection is stated using Bayes' theorem.

with respect to a minimal value in the affected area. Therefore, Grid Search is required for finding the optimal hyperparameter combination. 115 pixels for the SGD model's input train. Begin by configuring the loss function, maximum iteration periods, and starting learning rate. Then, apply grid search. Lastly, make predictions after training the model.

SGD algorithm has several advantages:1) It is simple enoughtoachievetheequation.Itonlyusesbasicmaththeorem suchasderivingthegradient.2)Itisanefficientmethodunlike SVMcostsalotoftime.3)ApartfromtraditionalSGD,random SGD performs be better if several locally optimum solutions exist. 4) The CPU's performance may be greatly enhanced by the small-batch SGD.

A. Models for Deep Learning
Deep neural networks (DNNs) are essentially just neural networks with many hidden layers, while neural networks themselves are extensions of perceptrons. The term "multi-layer Perceptron" (MLP) may describe a DNN [15]. There are three distinct kinds of DNN layers that are defined by their relative placement: input, hidden, and output.

(1) DNN. Tensorflow is used as the foundation for building Neural Networks in this study. To begin, one may construct a DNN model with three thick layers, each with a unique number of neurons. With 784 pixels for every letter, the input shape is 784. The data category is represented by the 24 neurons in the output layers. To get better results, you may tweak the settings such as activation function, loss, optimizer, and epochs. In this study, the default values for other adjustable factors, such as learning rate, were used. The issue of overfitting is investigated by building two models with distinct architectures.
CNN, the second. The exceptional performance of convolutional neural networks (CNNs) when fed images sets them apart from other types of neural networks. [16] The convolutional, pooling, and fully-connected layers are the fundamental components of a convolutional neural network (CNN). A large number of parameters are adjustable, much like DNN. This model is built using a Convolutional Neural Network architecture, which includes the following layers: 2 convolution, 2 batch normalization, 2 max pooling, 1 flatten, 1 dropout, and 2 dense. In the two layers of convolution, there are 75 neurons and 25 neurons, respectively. In the two layers of denseness, there are 512 neurons and 24 neurons. Except for the last dense layer, which uses softmax, all of the

activation functions in this model are relu. After 20 iterations of fitting the model, set the optimizer to Adam, metrics to accuracy, and categorical_crossentropy as the loss function. Learning rate and all other settings are set to default.

Results

TABLE I.The Results of Different Machine Learning Algorithmswith Original Data.

Model Name	Evaluation Metric			
	testaccuracys core	*testprecisio n score*	*testrecallscore*	*testf1sco re*
RFC	0.8161	0.80	0.81	0.80
KNN (K=1)	0.7817	0.8038	0.7817	0.7812
Gaussi anNB	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

TABLEII. The Results of Different Machine LearningAlgorithms with Data Processed by PCA.

ModelN ame	EvaluationMetric			
	testaccuracys core	*testprecisions core*	*testrecall score*	*testf1 score*
RFC	0.087	0.09	0.09	0.09
KNN (K=1)	0.8209	0.8402	0.8209	0.8225
Gaussia nNB	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 Algorithmswith Data Processed by ISOMAP

ModelN ame	Evaluation Metric			
	testaccuracys core	*test precisionscor e*	*testrecall score*	*testf1score*
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

Since J and Z need the movement of fingers, only 24 letters are learned. However, when J and Z are used for prediction, the outcome is determined by which letter is the most similar.

A. RFC outcome and analysis

Our results show that ISOMAP achieves better accuracy after dimensionality reduction in RFC compared to PCA. This is likely due to the fact that our training dataset does not contain a large number of linear relationships between pixels, and while both methods have their benefits, ISOMAP is more flexible in learning a broad class of nonlinear manifolds [17]. On the other hand, PCA isn't always superior than ISOMAP. The effectiveness of both linear and nonlinear methods for extracting features is investigated in [18]. Issues including sparsity, complexity, and noise are better handled by feature extraction algorithms (FEAs) in actual data sets [18]. Nonlinear FEA has been suggested as a solution to the problems with linear approaches, although studies have shown that it works better on human tasks than on real tasks [17]. Not to mention that nonlinear FEAs aren't always succeed where PCA fails[17] because to the possibility that they are vulnerable to dimensional curses. Because of their distinct dimensional curse sensitivity, PCA outperforms ISOMAP in terms of accuracy.

B. KNN research findings and analysis

Using KNN for data training and prediction, we discovered that a lower K yields better results. We selected K with the greatest accuracy via cross validation. Afterwards, we discovered that the accuracy was greater than previously when we utilized the data that had been dimensionality-reduced using PCA and ISOMAP. More than 10% improvement in accuracy was seen when ISOMAP was used.

Section C: GNB findings and analysis

The accuracy score rises by around 20% following PCA, as seen above, when compared to the original data. Regardless, GNB is not very effective in SLR. We are unaware of any extensive usage of GNB for pattern recognition; its primary use seems to be spam detection. After principal component analysis (PCA), the original data set is able to attain pixel independence and normal distribution. When you see the terms "price" or "discount" in an unsolicited email, it's probably spam. The independence requirement is also satisfied since the likelihood of such a word's occurrence has no effect on other words. The converse is also true; there is no plausible explanation for the idea that other people do not care about pixel values. Because principal component analysis (PCA) is a kind of linear projection technique, each pixel might be more or less independent of others after PCA, and it is possible that this conformity to the normal distribution of pixel values is thus unavoidable.

E. SVM outcome and analysis
Out of all the machine learning approaches, SVM is the most effective. Following principal component analysis, the accuracy score of the original data is quite close to the PCA accuracy value. However, ten minutes may be cut from the training period following PCA.
In order to determine the optimal hyperparameter values, we use Grid Search. As shown in Table IV below, we experimented with many potential values for the hyperparameters.

TABLEIV.HYPERPARAMETERSFORSVM.

Kernel	**HyperparameterCombination**		
	linear	***poly***	***rbf****
C	0.1	1*	10
Degree(when kernel=poly)	2	3	4

*meansthebestvalue

We find that the optimal combination is kernel=rbf with C=1 at the end.

A. Discussion and results for SGD
When it comes to original data and dimension-reduction data, SGD does not do very well. We chose a 'elastic net' penalty, a 'log' loss function, and a 'optimal' learning rate. Logistic regression is a probabilistic classifier that is obtained from the 'log' loss. The 'elasticnet' falls somewhere in the middle between ridge and lasso regression.

which make an effort to lessen the overfitting. The outcomes of the deep learning models are shown in Table V.

TABLEV.THERESULTSOFDEEPLEARNINGALGORITHMS.

Model Name	**EvaluationMetric**			
	loss	***accuracy***	***val_loss***	***val_accuracy***
DNN-1(1024, 512,24 neurons)	0.0525	0.9958	6.3743	0.8292
DNN-2(24, 24,24 neurons)	0.6890	0.7580	3.7469	0.4169
CNN	0.0014	0.9997	1.0108	0.9387
CNN afterAugmentation	0.0547	0.9854	0.0962	0.9781

ResultanddiscussionforNeuralNetworks

The neural network refrains from using the data during dimensionality reduction to guarantee the model's correctness and performance. The first DNN model discovered that the loss value was the opposite of what was expected and that the accuracy was much greater on the training set than on the test set. An indication of over-fitting is when the loss value keeps going up in this kind of neural network as shown in the training history analysis. Next, reduce the amount of neurons in each layer. Even now, overfitting is a concern. The accuracy in both the training set and the test set diminishes as the number of neurons decreases. The accuracy in the training set is 0.758, the accuracy in the test set is 0.4169, and the loss in the test set is 3.7469 when each layer has 24 neurons. Since DNN is a basic model, the results are unsatisfactory. Overfitting is simple to happen, yet it's reasonable to assume that accuracy and loss both improve with more neurons. With respect to performance, CNN outperforms DNN. The precision of the model is impacted by several factors. Following parameter tuning, accuracy increases to 0.9997 on the training set and 0.9387 on the test set. An improvement has been made to the issue of overfitting. Convolutional neural networks (CNNs) are superior than DNNs because, like human vision, they can efficiently compress high-dimensional pictures into smaller ones while preserving their unique features. Consistently better model performance follows Data Augmentation. There is a decrease in loss and an accuracy of 0.9781 in the test set.

Conclusion

The authors of this study suggested a system for sign language recognition that would use a variety of techniques to teach models to identify and categorize 24 different gesture letters. The letters J and Z are not included in the 26 since they need finger movement. Any letter that does not fit neatly into one of these 24 groups will be considered the closest match. In order to decrease data dimensionality and speed up training, this research created PCA and Manifold Learning. Next, we compare the results of RFC, KNN, GNB, SVM, and SGD. We train DNN and CNN models simultaneously to see how well they fare. The suggested approach was tested in a number of trials. The data set underwent a reduction from 784 to 115 dimensions, a process that is best executed by Manifold Learning. Not only can algorithms vary in their performance, but there is also variation in whether or not an algorithm is enhanced following dimensional reduction using various approaches. With a test accuracy of 0.8419, SVM performs best when working with raw data. The SVM has a positive impact with a value of 0.8515 after PCA dimensionality reduction. Nevertheless, KNN's performance was much enhanced, reaching 0.9654, after ISOMAP dimension reduction. The optimal hyperparameter combination for this research is.kernel=rbf and C=1, which were found using Grid Search. Neural Networks (CNNs) are powerful models in their own right; with data augmentation, CNN performance skyrocketed, and the test set accuracy ultimately reached 0.9781. After Data Augmentation, CNN performs better in a complete comparison. More research into applying other evaluation techniques and expanding the application of these models to more machine learning tasks is planned for the future.

References

[1] R. Rastogi, M. Shashank, and A. Sajan. "A novel approach forcommunication among Blind, Deaf and Dumb people." 2015 2ndInternational Conference on Computing for Sustainable GlobalDevelopment (INDIACom). IEEE, 2015.

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

[3] K. Agatonovic, R. Beresford, "Basic concepts of artificial neural network(ANN) modeling and its application in pharmaceutical research." Journalof pharmaceutical and biomedical analysis, vol. 22, pp. 717-727, 2000.

[4] Y.Qiu, etal."Pose-guidedmatchingbasedondeeplearningforassessingqualityofactiononrehabilitationtraining." BiomedicalSignalProcessingand Control, vol. 72, 103323, 2022.

[5] M. Sharma, P. Ranjna, and K. Ashok. "Indian sign language recognitionusing neural networks and KNN classifiers." ARPN Journal ofEngineering and Applied Sciences, vol. 9, 1255-1259, 2014.

[6] P. Haque, D. Badhon, and N. Nazmun. "Two-handed banglasignlanguagerecognitionusingprincipalcomponentanalysis(PCA)andKNNalgorithm." 2019 International Conference on Electrical, Computer andCommunication Engineering (ECCE). IEEE, 2019.

[7] J. Liu, J. Chen, and J. Cheng. "Static Gesture Recognition System forHuman-Computer Interaction." Infrared and Laser Engineering, 2002.

[8] Tecperson. "Sign Language MNIST." Kaggle,https://www.kaggle.com/datasets/datamunge/sign-language-mnist.

[9] Keras. "Keras: Deep learning library based on Python".https://keras.io/zh/preprocessing/image/

[10] "HowtoCreateaDecisionTreeClassifierinPythonUsingSklearn."Learning about Electronics,http://www.learningaboutelectronics.com/Articles/How-to-create-a-random-forest-classifier-Python-sklearn.php.

[11] "Whatisa KNN (K-Nearest Neighbors)",https://www.unite.ai/what-is-k-nearest-neighbors/

[12] H.Zhang,TheoptimalityofNaiveBayes."Proc.FLAIRS,2004.

[13] N. Amor, B. Salem, and E. Zied. "Naive Bayes vs decision trees inintrusion detection systems."Proceedings of the 2004 ACM symposiumon Applied computing, 2004.

[14] S. Karamizadeh, et al. "Advantage and drawback of support vectormachine 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 dimensionalityreduction for visualization and classification."IEEE Transactions onSystems, Man, and Cybernetics, Part B (Cybernetics), vol.35, pp.1098-1107,2005.

[18] F. Anowar, S. Samira, and S. Bassant. "Conceptual and empiricalcomparison of dimensionality reduction algorithms (pca, kpca, lda, mds,svd, lle, isomap, le, ica, t-sne)."Computer Science Review, vol.40, 2021.