Ensemble Learning based Hand Gesture Recognition using Deep Convolutional Characteristics from Pre-trained CNN

Sunil G. Deshmukh[a]*, Shekhar M. Jagade[b]

[a]Department of Electronics and Computer Engineering, Maharashtra Institute of Technology,
Aurangabad, Maharshtra-431010, India
[b]Department of Electronics and Telecommunication, N B Navale Sinhgad College of
Engineering, Solapur, Maharashtra-413255, India

[*]**Corresponding author:** Sunil G. Deshmukh, **Email:** sunildeshmukh7472@gmail.com

Abstract

The variety and ambiguity of the hand gestures will have a significant impact on the accuracy and trustworthiness of the recognition task. In the modern age, hand gesture recognition has replaced human interactions. However, in the fields of computer imagination and pattern recognition, detecting the hand component has grown to be a difficult problem. The manual feature extraction approach used in the past is laborious and has a poor recognition rate when it comes to recognizing hand gestures. A new recognition method relying on convolutional neural network (CNN) is suggested in order to increase the recognition rate. Convolutional neural network (CNN) architectures are a highly popular deep learning approach for classification applications, but they have certain drawbacks, including large variation during forecasting, overfitting issues, and prediction mistakes. Nevertheless, proper training a CNN model takes a lot of time. This work presents an ensemble of CNN-based methods to address these issues. A graded ensemble model is provided, which boosts overall network functionality by using the supplementary data supplied by the base model. In this study, deep convolutional feature extraction is carried out using a pre-trained CNN model. Also, the ensembles of machine learning (ML) model are employed as a light weighted modeling over deeper learning (DL) approaches in relationships of lowering the training expense for the image classifications problem employing deep convolutional features. The dataset is used to test the suggested ensemble technique, and the accuracy attained is 99.1%. It has been shown that our suggested ensemble model accomplishes better than other suggested techniques already in use.

Keywords: CNN, Hand gesture recognition, Deep convolutional characterstics, VGG-16, VGG-19, Xception Conv block.

1. Introduction

The frequency of interaction between people and computers has significantly increased, and the field is always growing as new methods and algorithms are developed. One of the most difficult areas where technologies has evolved is hand gesture detection. Gesture-based signalling systems have been complemented by artificial intellect and computer recognition, which have enhanced communications among the deaf [1, 2]. A few of the subdivisions of hand gesture recognition include sign language interpretation [3-5], human action interpretation [6], and position and posture detection [7, 8]. A common kind of nonverbal communication is hand motions. It is made up of sign language linguistic expertise that contains a lot of relevant data. Systems for hand gesture recognition are essential components of human-computer interpretation (HCI) systems. Thus, there is a high need for algorithmic hand gesture implementations. Since the end of the previous century, this area has attracted the interest of several scholars. The hand gesture recognition technique has drawn attention largely due to the following factors [9]. (i) The sharp rise in the number of hearing-impaired persons, and (ii) the widespread usage of vision-based and touchless platforms and gadgets. Strong hand gesture recognition is an essential component of sign language interpreting for hearing-impaired people. The individuals with normal hearing and those with hearing loss have noticeable communication gaps. A translation strategy that converts the gestural languages into idiomatic phrases may close this gap in communication. Implementing an interpretation system may help hearing-impaired persons, allowing them to more easily and independently integrate into community.

The proper identification and classification of hand gestures is the main goal of the system for recognizing hand gestures. In order to understand the forms and postures of the hands, hand recognizing uses a variety of methods and concepts from several fields, including image processing and machine learning. Accuracy rate should be the end objective of a dependable recognition program, and convolutional neural networks, one kind of deep neural network, excel at recognizing complicated patterns. This research introduces an ensemble method for transfer learning-based hand posture identification.Due to its capacity to automatically retrieve deep convolution information, deep learning (DL) frameworks like convolutional neural networks (CNNs) are often taken into consideration among AI-based algorithms. Nevertheless, training a CNN model takes a lot of time. As a result, the effectiveness of the models heavily depends on the conventional feature extraction techniques. However, manually extracting features takes a lot of effort and is prone to mistakes. Since they can automatically extract important information, deep learning (DL) models like CNNs have surpassed more conventional AI models as a preferable option.

A CNN model is composed of two components: feed-forward neural networks for categorization and convolutional modules for deep convolutional extracting features [10]. The training procedure for CNN models, which include millions of trainable features, takes a very long period. As a consequence of this, the concept of transfer learning is used in this investigation via the utilization of a pre-trained model deep CNN (DCNN) modelling for the deep neural layer gathering features in an effort to shorten the training timescales of a newly constructed CNN model. In this study, the latent representation of the inputs or the flattened vector is regarded as the deep convolutional characteristic. The categorization of hand movements based on deep convolutional aspects is then performed using an ensemble of five machine learning models. By replacing the final layer neural network of the DCNN model with a lighter alternative, Training time may be cut down significantly by using low-weight machine learning models to solve the classification issue. The machine learning models that are taken into consideration by the ensemble methods include the k-nearest neighbours (KNN), multi-class linear regression (LR), random forest (RF), extreme gradient boosting (XGBoost), and support vector machine (SVM) [11]. Also, during the preparatory phase, the input hand motion photographs go through stages of interference elimination, image enrichment, and segmentation techniques. The most important detail from the incoming images is extracted using this three-step pre-processing approach.

The Kaggle gesture recognition datasets are used to test the recommended ensemble learning-based hand gesture categorization model. The dataset consists of 7,172 testing images and 27,455 training visuals 5000 samples. There are a total of 10 classes that are thought to comprise 5 different hand gestures apiece. 99.1% accuracy was attained using the suggested model. In this paper, a comparison of three pre-trained DCNN modeling techniques, the

VGG-19, Xception model's, VGG-16and intense convolutional features is also provided in terms of classification accuracy. The remainder of the essay is structured as follows. The summary of relevant publications is included in Section 2 literature review. The suggested strategy and the pre-processing methods are covered in section 3. Section 4 includes a description of the datasets and the experimental setup. The section 5 report includes both the outcomes and the perspectives that pertain to the assessment methods. In the 6th section, the conclusion as well as the works should be performed as futuristic efforts has been discussed.

2. Literature review

Several conventional machine learning (ML) techniques were used in the early stages of the AI-based automatic hand gesture identification. There is many research on gesture recognition available since the field of hand gesture recognition is rapidly increasing. Here, some of the key ideas and strategies are covered. Flores et al. [12] offer a static hand gesture recognition model with persistent characteristics using two CNN networks. The accuracies achieved by the researchers were 95.37% and 96.20%, respectively. Alani et al. [13] published proposal for an adjusted deep convolutional neural network incorporated CNN and data pre-processing methods. For validation, the authors employed a dataset of 3750 static hand gesture photos. The ADCNN in comparison to a CNN mode was used by the authors to achieve a recognition accuracy of 99.73%. A CNN model was suggested by Han et al. [14] for the identification of static hand gestures. A collection of 12,000 pictures of 10 different hand motions was used by the authors. The train and test datasets are obtained during the picture pre-processing stage using the Gaussian skin modeling and background suppression. A simple six-layered CNN created by the authors has an overall categorization accuracy of 93.8%. Ameen et al. proposed a CNN that can categorize both depth and color pictures [15]. For the ASL dataset, the authors' accuracy and recall rates were 82% and 80%, respectively. Moreover, Chong et al. [16] have created a model based on the jump motion controller. For recognition, the whole set of ASL's 26 letters and 10 numbers has been utilised. Using a supported vector algorithm and a deep neural network, the authors' identification rates were 80.3% and 93.81%, respectively.

An effective Convolution neural based static sign linguistic recognition system has been presented by Wadhan et al. [17]. The efficacy of the network was assessed using fifty different CNN models, according to the authors. There have been 100 static indicators from different sources considered. For the grayscale and color pictures, respectively, training accuracy of 99.90% and 99.72% have been reached. Barbhuiya et al. [18] have created a thorough neural learning system for sign language recognition. Before utilizing a binary classification SVM decoder for classifying, features were extracted using a tailored version of AlexNet and the VGG16 network. A maximum recognition accuracy of 99.82% was reached by the authors. For domain adaptation, Can et al. [19] worked in deep learning models: VGG16, VGG19 and hand gesture images have both been taken into consideration for identification. Employing near-infrared pictures, the highest identification accuracy of 100 percent has been attained. Tan et al. [20] proposed a multilayer neural networks with spatial pyramids pooling for gestures identification. Even with inputs of various sizes, spatial pyramid pooling generates a fixed-length feature representation. A peak recognition rate having accuracy of 99.03% was reached by the authors. A traditional and deep learning-based proposed feature selection model for gesture identification was also presented by Barbhuiya et al. [21]. To obtain the local metadata of the picture, the scientists took into account local as well as global image aspects. 99.84% accuracy in maximum determination has been attained.

In conclusion, the major objective is to provide a framework for absolute hand gesture detection that overcomes the most frequent problems while lowering constraints and producing accurate results. Limitations in the accurate detection and categorization of hand postures include backdrop difficulties, similarity across classes, lighting variations, and distance range. In order to improve the cumulative recognition accuracy and resilience of the model, we have created an ensemble model in this study that makes advantage of the complementing information provided by the various base modes. As a result of the diversity of the information, ensemble learning outperforms individual models. When the verdict of many models is taken into account, less noisy forecasts are produced. As a result, this strategy has been used in the current work. We also demonstrate how poorly supervised segmentation may be carried out using the suggested approach.

3. Methodology

This research offers a reliable and effective method for categorizing and identifying intricate ASL hand movements. To reduce the false positives and false negatives with limited computing resources, deep ensemble neural network with transferring pedagogy is presented. Fig. 1 shows the recommended methodology's architectural layout. In this research, ensemble learning is implemented using pre-trained multilayer network models such the VGG-16 [22], VGG-19 [23], and Xception [24] models.

In this study, deep convolutional characteristics are extracted from the segmentation pictures using three pre-trained DCNN frameworks. The DCNN models under consideration include VGG-19, Xception model's, VGG-16. Nevertheless, since the suggested strategy uses pre-trained DCNN models, the trainable attributes evaluated in this study are zero. However, for the VGG-19, Xception model's, VGG-16 algorithms, the underlying encoding of the input pictures has a size of 18191, 18191, and 32768, respectively. The ensemble network of 5 multi-layer models depicted in Fig. 1 is then trained using latent vectors belonging to each prediction.

XGBoost, KNN, RF, SVM, and multi-class LRmake the recommended ensemble learning-based hand gesture identification technique. The complex convolutional properties that were retrieved utilizing the taken into consideration DCNN models were first used to educate and validate each unique ML model. Relying on the results from each model's individual tests, it was discovered that SVM and sub LR models had somewhat higher accuracy than KNN, RF, and XGBoost. Also, it was discovered that VGG-16-based fully convolutional features outperformed the competition among the DCNN models that were taken into consideration. As a result, VGG-16 based deep convolutional features are taken into account in the final ensemble model.

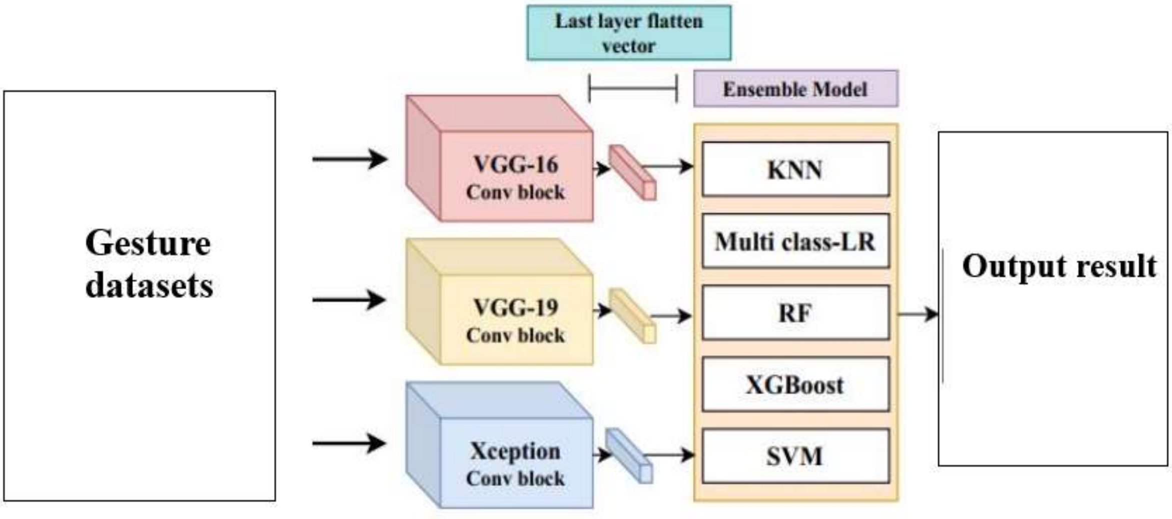

Fig. 1. Deep ensemble neural network methodology's architectural layout

Utilizing deep convolutional layers from the VGG-16, VGG-19, and Xception models into the ensemble's modeling strategy, accuracy was reached at 99.1%, 96.7%, and 98.9%,correspondingly.

3.1.VGG16 and VGG19

As its name implies, VGG16 is a 16-layer deep neural network. The VGG16 network is thus fairly huge by today's standards with 138 million total variables [22]. The VGGNet design incorporates the key elements of convolutional neural networks [23]. The pre-trained VGG16 networks supervised on the ImageNet dataset has been considered for transfer learning. Instead of creating a new model, transfer learning is an easy and affordable method. as building a new network from scratch requires a lot of time and resources. The retrieved characteristics from the locally connected layer and VGG16 architecture are multiplied to create an attention module. Rectified linear unit (ReLU), which gives the same result for positives and negatives input data, yields zero, serves as the activation function in this case. A VGG network is composed of tiny convolution filters. 13 convolutional layers and 3 fully connected layers make up VGG16. Fig. 2 below gives a general overview of the VGG architecture for the highlighted extraction. 16 convolutional layers and 3 fully linked layers make up the 19 layers that make up VGG-19. Compared to the VGG-16, the VGG-19 has more layers and a deeper CNN framework.

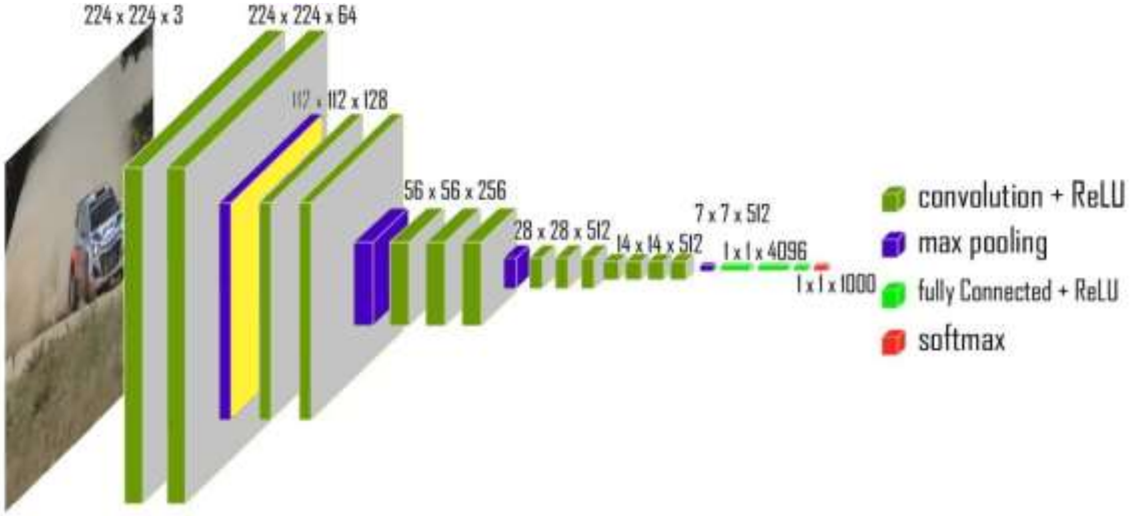

Fig. 2. VGG architecture model

3.2 Xception model

As a CNN framework, the Xception model was suggested. "Extreme inception" was the name given to it. Xception features 36 layers of convolutions. There are three flows in it. The entrance flow, which is the initial flow, contains layers for convolution, separable convolution, and pooling. The intermediate flow, which contains separable convolution layers, is the second flow. Eight times the centre flow is repeated. The exit flow makes up the third flow. It is the final phase of flow and produces a thick layer as a consequence. Fig. 3. depicts the model of Xception architecture. The Xception method was chosen since it is very efficient and produces excellent outcomes in prior studies [25, 26].

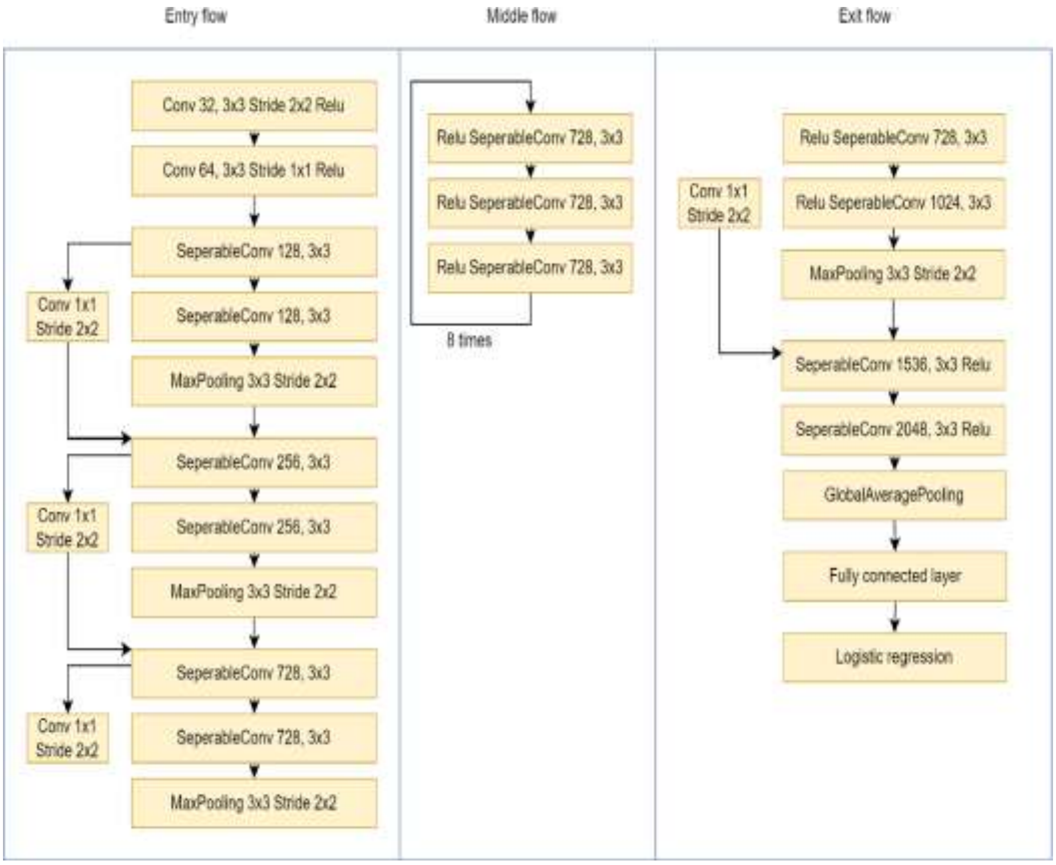
Entry flow
Middle flow
Exit flow
Conv 32, 3x3 Stride 2x2 Relu
Conv 64, 3x3 Stride 1x1 Relu
SeperableConv 128, 3x3
Conv 1x1 Stride 2x2
SeperableConv 128, 3x3
MaxPooling 3x3 Stride 2x2
SeperableConv 256, 3x3
Conv 1x1 Stride 2x2
SeperableConv 256, 3x3
MaxPooling 3x3 Stride 2x2
SeperableConv 728, 3x3
Conv 1x1 Stride 2x2
SeperableConv 728, 3x3
MaxPooling 3x3 Stride 2x2
Relu SeperableConv 728, 3x3
Relu SeperableConv 728, 3x3
Relu SeperableConv 728, 3x3
8 times
Relu SeperableConv 728, 3x3
Conv 1x1 Stride 2x2
Relu SeperableConv 1024, 3x3
MaxPooling 3x3 Stride 2x2
SeperableConv 1536, 3x3 Relu
SeperableConv 2048, 3x3 Relu
GlobalAveragePooling
Fully connected layer
Logistic regression

Fig. 3. Xception architecture model

3.3.Ensemble model

Using a variety of learning models in combination, ensemble learning is a well-liked machine learning approach that improves model accuracy and lowers prediction inaccuracies [27]. It is renowned for being very reliable and accurate than any specific CNN model. As the synchronous ensemble model offers both sequential and parallel approaches, we have employed it in our investigation. By aggregating the final results of these 3 divergent models, the approach employs three separate models running simultaneously. The resulting gesture description is then predicted using this average score.

4. Experimental environment

Python 3.6.9. software is used for the execution of proposed model of ensemble based deep learning approach. The details of the system description that have been used for experimentation are as follows; Dell Precision 7920: 32GB RAM, 500GB HDD,Intel XeonBronze 3204, 8.25MB cache, 6 cores, 6 threads, 1.90 GHz, AMD Radeon Pro W6400, 4GB, 2DP (Precision 7920T).

4.1.Description of the datasets

Data collecting is one of the most important components of any inquiry. It is essential to collect data that is relevant to the research and meets with its requirements. The suggested research project's dataset was created using the Kaggledataset. The classic Kaggle dataset contains of transcribed letters is a well-known landmark for visual machine learning methods. Nonetheless, Researchers have intensified their attempts to modernize it and create drop-in alternatives that are unique for use in the physical world and more difficult for machine learning [20].

The dataset consists of 7987 testing images and 31948 training visuals. Each image is 28 by 28 pixels in size and is composed entirely of grayscale. Each image is labelled with the letter of the alphabet it corresponds to. The letters "j" and "z," which both suggest motion, have been excluded from this dataset's 25 letters. Details of the dataset labels is shown in Table 1.

Table 1. Details of dataset labels

Alphabet Name	Label
A	0
B	1
C	2
D	3
E	4

F	**5**
G	**6**
H	**7**
I	**8**
J	**9**
K	**10**
L	**11**
M	**12**
N	**13**
O	**14**
P	**15**
Q	**16**
R	**17**
S	**18**
T	**19**
U	**20**
V	**21**
W	**22**
X	**23**
Y	**24**
Z	**25**

5. Result and Discussion

The first step is to individually train and evaluate each of the five models (XGBoost, KNN, RF, SVM, and multi-class LR) using a pre-processed datasets of hand gestures. Using 4000 and 1000 data inputs, respectively, that include deep convolutional layers matching to each picture, each model is given training and evaluated. For all of the DCNN model types, the deep convolutional properties are taken into account (i.e., VGG-19, Xception model, and VGG-16).

Findings are initially gathered for a test dataset made up of deep convolutional attributes based on the Xception framework. Table 2.displays the testing accuracy results for each particular model. The precision, recall, F1 score, and accuracy of each model under consideration are listed in Table 2. It can be shown that multiclass LR, XGBoost, and SVM exhibit adequately good performance amongst the ML models taken into consideration. The ensemble model's accuracy for the testing data is then determined. The ensemble model has the best accuracy (98.9%), as shown in Table 2, when compared to the other models that were taken into consideration. It suggests that the suggested ensemble learning strategy is better than the other individual ML models that have been taken into consideration.

A method for quantifying a classifying algorithm's effectiveness is the confusion matrix. For the hand gesture given datasets, a confusion matrix is used to assess the effectiveness of a classifier's output. The value of data sites for which the anticipated label matches the actual label is shown by the diagonal matrix, whilst the incorrect labels assigned by the classifier are represented by the off-diagonal attributes. The confusion matrix's diagonal elements should be high, suggesting numerous accurate predictions. In Fig. 4, the suggested ensemble method for deep convolutional features based on the Xception model corresponds confusion matrix. The model's efficacy is then calculated on the testing datasets, which contains deep convolutional features based on the VGG-19 model.

Table 3 displays the testing accuracy for each unique model. Table 3 makes it evident that multi-class LR and SVM perform much better than the respective individual models. Nonetheless, Table 3 demonstrates that the ensemble model's accuracy is the greatest among the three individual estimates at 96.7%. In Fig. 5, the suggested ensemble classifier for deep convolutional attributes based on the VGG-19 is shown with its confusion matrix. Lastly, the performance of the deep convolutional layers based on the VGG-16 prototype is evaluated. With the VGG-16 based deep convolutional features, the test effectiveness of each individual model is first calculated in a similar way. Table 4 displays the models' test accuracies. As can be seen in Table 4, each model's performance significantly improved when comparing to deep convolutional feature architectures based on the VGG-19 and Xception algorithms. This shows that when contrasted to certain other DCNN models, deep convolutional layers based on VGG-16 are significantly better at classifying hand gestures. Moreover, Table 4 shows that multiclass LR, XGBoost, and SVM perform with performance of 96.6%, 91.4%, and92%, respectively, which is considerably superior than other investigated models. While multi-class LR has an accuracy of 96.6%, it can be shown from Table 4 that the recommended ensemble training model performs the best in regard to precision (i.e., 99.1%). This suggests that our suggested ensemble learning-based gesture categorization model outperformed all other distinct machine learning models. In Fig. 6, the suggested ensemble method for VGG-19-based deep multilayer features corresponds confusion matrix. The label accuracy, recall, and F1 ratings are also shown in Table 5 in order to demonstrate how well the suggested ensemble learning strategy performs with respect to the deep convolutional models based on the VGG-16. The suggested model performed adequately for each specific class, as shown in Table 5. This suggests that the ensemble pedagogical strategy is successful for the job of gesture categorization.

Table 2. Class labels for Xception model.

Models (M)	**Precision**	**Recall**	**F1 Score**	**Accuracy in %**
KNN	0.864	0.864	0.864	86.4
Muti-Class LR	0.942	0.942	0.942	94.2
RF	0.885	0.885	0.885	88.5
XGBoost	0.913	0.913	0.913	91.3
SVM	0.928	0.928	0.928	92.8
Ensemble	**0.989**	**0.988**	**0.989**	**98.9**

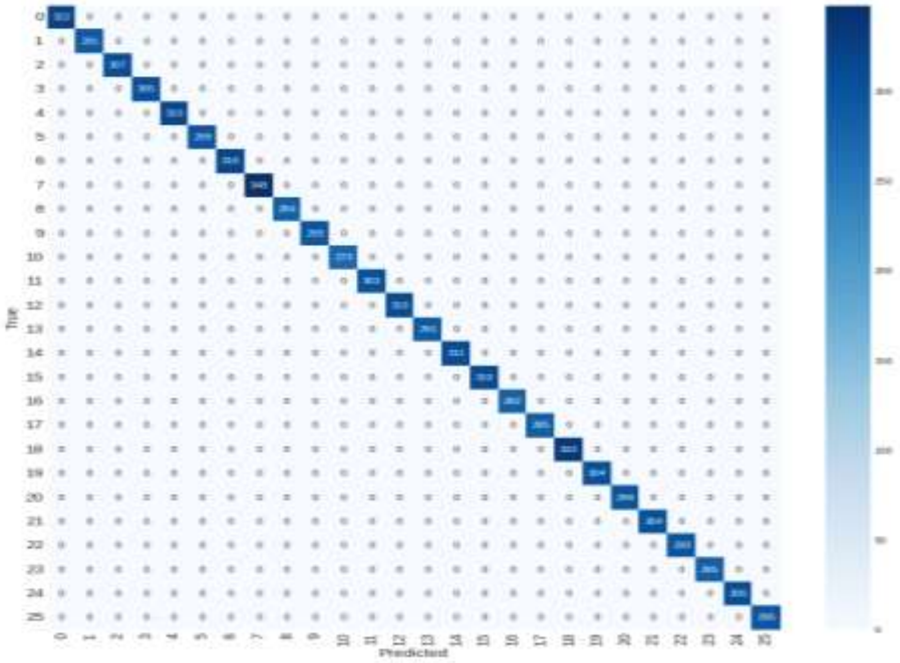

Fig. 4. Confusion matrix for Xception model with ensemble based deep convolutional

Table 3. Class labels for ensemble of VGG-19

Models (M)	Precision	Recall	F1 Score	Accuracy in %
KNN	0.814	0.814	0.814	81.4
Muti-Class LR	0.942	0.942	0.942	94.2
RF	0.877	0.877	0.877	87.7
XGBoost	0.905	0.905	0.905	90.5
SVM	0.951	0.951	0.951	95.1
Ensemble	**0.967**	**0.967**	**0.967**	**96.7**

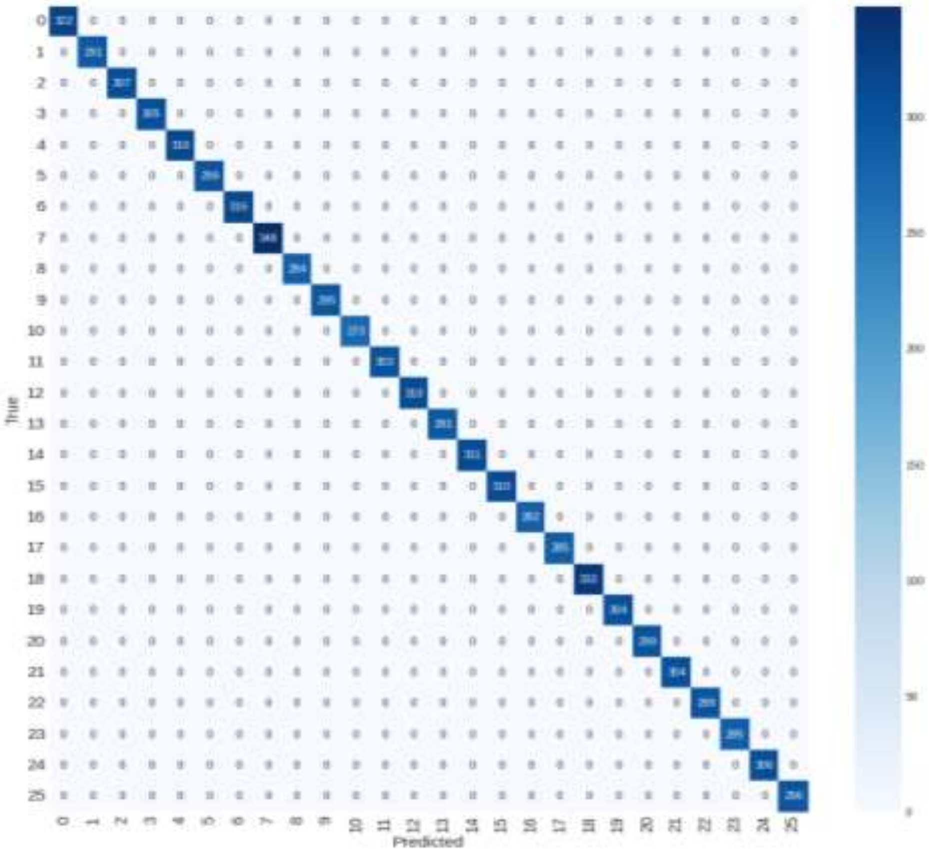

Fig. 5.Confusion matrix for VGG-19 framework having ensemble classifier.

Table 4. Class labels for ensemble of VGG-16

Models (M)	Precision	Recall	F1 Score	Accuracy in %
KNN	0.854	0.854	0.854	85.4
Muti-Class LR	0.966	0.966	0.966	96.6

RF	0.867	0.867	0.867	86.7
XGBoost	0.914	0.914	0.914	91.4
SVM	0.92	0.92	0.92	92.0
Ensemble	**0.991**	**0.991**	**0.991**	**99.1**

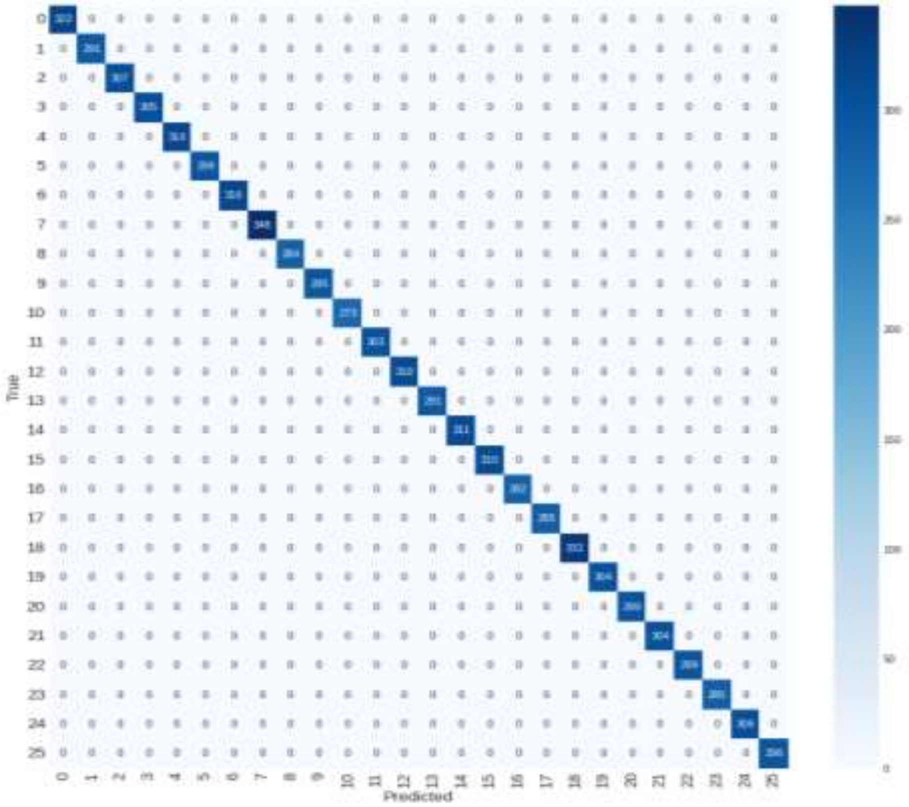

Fig. 6. Confusion matrix for VGG-16 framework having ensemble classifier

Table 5. Class labels ofVGG-16 method

Class	**Precision**	**Recall**	**F1 Score**
A	0.96	0.94	0.96
B	0.97	0.97	0.97
C	0.98	0.95	0.96
D	0.98	0.96	0.94
E	0.96	0.95	0.95
F	0.97	0.97	0.97
G	0.96	0.96	0.96

H	0.99	0.97	0.98
I	0.99	0.98	0.98
J	0.98	0.95	0.96
K	0.98	0.97	0.97
L	0.97	0.97	0.97
M	0.97	0.96	0.96
N	0.96	0.96	0.96
O	0.97	0.97	0.97
P	0.98	0.95	0.98
Q	0.98	0.96	0.95
R	0.97	0.95	0.97
S	0.96	0.96	0.96
T	0.99	0.98	0.96
U	0.96	0.96	0.96
V	0.98	0.98	0.98
W	0.99	0.96	0.95
X	0.99	0.99	0.98
Y	0.97	0.98	0.98
Z	0.98	0.98	0.98

6. Conclusion

An ensemble learning strategy is suggested in this work for the job of classifying hand gestures. In this study, the ML models of XGBoost, KNN, RF, SVM, and multi-class LR are taken into consideration. A pre-trained DCNN model is employed as a deep convolutional featured extractor in this study, which also incorporates a novel feature extraction technique. In order to extract the important information from the input data, a 3-phase visual pre-processing approach is also used in this study. This shows that when comparing with other DCNN models, deep convolutional aspects based on VGG-16 are significantly superior at classifying hand gestures. Furthermore, Table 5 shows that multiclass LR, XGBoost, and SVM perform with performances having 96.6%, 91.4%, and 92%, correspondingly, which is considerably better than those certain investigated models. Despite the fact that multi-class

LR has an accuracy of 95.3%, it can be inferred from Table 5 that the suggested ensembles learning-based model performs the better in regards of accuracy (i.e., 99.1%). This suggests that our suggested ensemble learning-based gesture categorization model outperformed all other individual machine learning models. In order to improve the performance of the models, there is potential to use meta-heuristic optimizing strategies for deep fully convolutional investigation in future findings.

The Kaggle hand gesture datasets are available at;

https://www.kaggle.com/datasets/grassknoted/asl-alphabet

References

1. Fang, Y.; Wang, K.; Cheng, J.; Lu, H.: A real-time hand gesture recognition method. In: 2007 IEEE International Conference on Multimedia and Expo, pp. 995–998. IEEE (2007)
2. Oudah, M.; Al-Naji, A.; Chahl, J.: Hand gesture recognition based on computer vision: a review of techniques. J. Imaging **6**(8), 73 (2020)
3. Al-Hammadi, M.; Muhammad, G.; Abdul, W.; Alsulaiman, M.; Bencherif, M.A.; Alrayes, T.S.; Mathkour, H.; Mekhtiche, M.A.: Deep learning-based approach for sign language gesture recognition with efficient hand gesture representation. IEEE Access **8**, 192527–192542 (2020)
4. Vaitkeviˇ cius, A.; Taroza, M.; Blažauskas, T.; Damaševiˇ cius, R.; Maskeli¯ unas, R.; Wo´ zniak, M.: Recognition of American sign language gestures in a virtual reality using leap motion. Appl. Sci. **9**(3), 445 (2019)
5. Rezende, T.M.; Almeida, S.G.M.; Guimarães, F.G.: Development and validation of a Brazilian sign language database for human gesture recognition. Neural Comput. Appl. **33**(16), 10449–10467 (2021)
6. Afza, F.; Khan, M.A.; Sharif, M.; Kadry, S.; Manogaran, G.; Saba, T.; Ashraf, I.; Damaševiˇ cius, R.: A framework of human action recognition using length control features fusion and weighted entropy-variances basedfeatureselection.ImageVis. Comput. **106**, 104090 (2021)
7. Nikolaidis, A.; Pitas, I.: Facial feature extraction and pose determination. Pattern Recogn. **33**(11), 1783–1791 (2000)
8. Kulikajevas, A.; Maskeliunas, R.; Damaševiˇ cius, R.: Detection of sitting posture using hierarchical image composition and deep learning. PeerJComput. Sci. **7**, e442 (2021)

9. Kausar, S.; Javed, M.Y.: A survey on sign language recognition. In: 2011 Frontiers of Information Technology, pp. 95–98. IEEE (2011)
10. S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6.
11. S. Badillo, B. Banfai, F. Birzele, I. I. Davydov, L. Hutchinson, T. Kam- ´ Thong, J. Siebourg-Polster, B. Steiert, and J. D. Zhang, "An introduction to machine learning," *Clinical Pharmacology and Therapeutics*, vol. 107, pp. 871 – 885, 2020.
12. Flores, C.J.L.; Cutipa, A.G.; Enciso, R.L.: Application of convolutional neural networks for static hand gestures recognition under different invariant features. In: 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), pp. 1–4. IEEE (2017).
13. Alani, A.A.; Cosma, G.; Taherkhani, A.; McGinnity, T.M.: Hand gesture recognition using an adapted convolutional neural network with data augmentation. In: 2018 4th International Conference on Information Management (ICIM), pp. 5–12. IEEE (2018).
14. Han, M.; Chen, J.; Li, L.; Chang, Y.: Visual hand gesture recognition with convolution neural network. In: 2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), pp. 287–291. IEEE (2016).
15. Ameen, S.; Vadera, S.: A convolutional neural network to classify American sign language fingerspelling from depth and colour images. Expert. Syst. **34**(3), e12197 (2017)
16. Chong, T.W.; Lee, B.G.: American sign language recognition using leap motion controller with machine learning approach. Sensors **18**(10), 3554 (2018).
17. Wadhawan, A.; Kumar, P.: Deep learning-based sign language recognition system for static signs. Neural Comput. Appl. **32**(12), 7957–7968 (2020).
18. Barbhuiya, A.A.; Karsh, R.K.; Jain, R.: CNN based feature extraction and classification for sign language. Multimed. Tools App. **80**(2), 3051–3069 (2021).
19. Can, C.; Kaya, Y.; Kılıç, F.: A deep convolutional neural network model for hand gesture recognition in 2D near-infrared images. Biomed. Phys. Eng. Express **7**(5), 055005 (2021).
20. Tan, Y.S.; Lim, K.M.; Tee, C.; Lee, C.P.; Low, C.Y.: Convolutional neural network with spatial pyramid pooling for hand gesture recognition. Neural Comput. Appl. **33**(10), 5339–5351 (2021).

21. Barbhuiya, A.A.; Karsh, R.K.; Jain, R.: A convolutional neural network and classical moments-based feature fusion model for gesture recognition. Multimed. Syst. **28**, 1779–1792 (2022).
22. S. Tammina, “Transfer learning using vgg-16 with deep convolutional neural network for classifying images,” *International Journal of Scientific and Research Publications (IJSRP)*, 2019.
23. J. Z. X. G. S. Xiao, J. Wang, S. Cao, and B. Li, “Application of a novel and improved vgg-19 network in the detection of workers wearing masks,” *Journal of Physics. Conference Series*, vol. 1518, 2020.
24. W. W. Lo, X. Yang, and Y. Wang, “An xception convolutional neural network for malware classification with transfer learning,” *2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, pp. 1–5, 2019.
25. Chen, B., Ju, X., Xiao, B., Ding, W., Zheng, Y., & de Albuquerque, V. H. C. (2021). Locally GAN-generated face detection based on an improved Xception. *Information Sciences,572*, 16–28.
26. Shaheed, K., Mao, A., Qureshi, I., Kumar, M., Hussain, S., Ullah, I., & Zhang, X. (2022). DS-CNN: A pre-trained Xception model based on depth-wise separable convolutional neural network for finger vein recognition. *Expert Systems with Applications,191*, 116288.
27. Ewe, E.L.R., Lee, C.P., Kwek, L.C. and Lim, K.M., 2022. Hand Gesture Recognition via Lightweight VGG16 and Ensemble Classifier. *Applied Sciences*, *12*(15), p.7643.