IJIST P ISSN : 2618-1630 E ISSN : 2909-6130 Recognized Journal

Original Article

Interpretation of Expressions through Hand Signs Using Deep Learning Techniques

Sameena Javaid[1*], Safdar Rizvi[1], Muhammad Talha Ubaid[2], Abdou Darboe[3], Shakir Mahmood Mayo[4]

[1]Department of Computer Sciences, School of Engineering and Applied Sciences, Bahria University Karachi Campus, Karachi 75290, Pakistan.

[2]National Center of Artificial Intelligence, KICS, University of Engineering and Technology, Lahore 39161, Pakistan.

[3]University of The Gambia, Serrekunda, The Gambia.

[4]University of Engineering & Technology Lahore

*Correspondence: Sameena Javaid (sameenajaved.bukc@bahria.edu.pk).

Citation | Sameena Javaid, Safdar Rizvi, Muhammad Talha Ubaid, Abdou Darboe, Shakir. Mahmood. Mayo "Interpretation of Expressions through Hand Signs Using Deep Learning Techniques," Int. J. Innov. Sci. Technol., vol. 4, no. 2, pp. 596–611, 2022.

Received | May 28, 2022; **Revised** | June 22, 2022; **Accepted** | June 23, 2022; **Published** | June 25, 2022.

DOI:https://doi.org/10.33411/IJIST/2022040225

It is a challenging task to interpret sign language automatically, as it comprises high-level vision features to accurately understand and interpret the meaning of the signer or vice versa. In the current study, we automatically distinguish hand signs and classify seven basic gestures representing symbolic emotions or expressions like happy, sad, neutral, disgust, scared, anger, and surprise. Convolutional Neural Network is a famous method for classifications using vision-based deep learning; here in the current study, proposed transfer learning using a well-known architecture of VGG16 to speed up the convergence and improve accuracy by using pre-trained weights. We obtained a high accuracy of 99.98% of the proposed architecture with a minimal and low-quality data set of 455 images collected by 65 individuals for seven hand gesture classes. Further, compared the performance of VGG16 architecture with two different optimizers, SGD, and Adam, along with some more architectures of AlexNet, LeNet05, and ResNet50.

Keywords: Pakistan Sign Language, Hand Gestures, Convolutional Neural Network (CNN), VGG16, Transfer Learning.

Acknowledgment. We confirm that this work is original and has not been previously published in whole or in part, and it is not being considered for publication elsewhere. **Project details.** The authors received no specific funding for this study.

CONFLICT OF INTEREST: We have no conflicts of interest to disclose, all authors have read the final manuscript and have approved the submission to the journal, and have accepted the full responsibility for the manuscript's delivery and contents. There exists no conflict of interest for publishing this manuscript in IJIST.

Author's Contribution. Conceptualization, S.J., S.R., and M.T.U.; Methodology, S.J., and S.R.; Software, S.J.; Validation, S.J., and M.T.U.; Formal Analysis, S.J.; Investigation, S.J. and M.T.U.; Resources, S.J., and S.R.; Data Curation, S.J.; Writing—original draft preparation, S.J.; Writing—review and editing, S.R.; Visualization, M. T. U.; Supervision, S.R.

INTRODUCTION

It is estimated that around 480 million people have hearing loss worldwide [1]. Sign Language (SL) is the way to reduce the gap between deaf and ordinary people. But there is the fact that several people with hearing loss even know sign language depends on translators or interpreters because many normal ordinaries are unaware of sign language. In this context, any system that automatically translates sign language is beneficial improving mute people's lives by increasing social enclosure and self-freedom of expression [2].

Consequently, developing technologies are improving the hurdles of deaf and mute people, but the problem sphere of sign language is still a challenging domain [3]. Among several problems, one of them is a variety of sign languages, just like spoken languages. Over the world, there is no global sing language. Therefore, to the best of our knowledge, the applicant converts a generic sign language interpreter to lessen the gap between mute, hard-of-hearing people and everyday individuals [4]. American Sign Language (ASL), Chinese Sign Language (CSL), Indian Sign Language (ISL), British Sign Language (BSL), etc., all are different. Subsequently, Pakistan Sign Language (PSL) has its rare as Urdu Sign Language (USL). Research and development space for Pakistan Sign Language is enormous; compared to other languages, PSL needs special attention to automate the physical characteristics of users and interpret them into meaningful text or conversation [5].

Understanding emotions in human gestures are essential to understanding subjective, physiological, and expressive components. These three elements play an essential role in understanding the emotional response of individual components, that how any individual experiences any emotion which creates the body's reaction and behavior, which are called physiological and expressive components, respectively [6].

With the development and growth in Machine Learning (ML), many applications benefit from interpreting Sign Language. Among several novel architectures of Machine Learning, Neural Network (NN) is one of the most commonly used architectures [7], [8]. Artificial Neural Network (ANN) simulates the behavior of biological neurons) it is built up of several layers of artificial neurons. Different features are detected by several specialized and dedicated layers of neurons. Since the signs in sign language are varied and consist of different patterns like hand posture, shifting, and movement, using an artificial neural network helps develop the systems that detect and interpret signs [9].

Recently, Convolutional Neural networks (CNN) have been growing fast for vision-based deep learning. Generally, when initial weights are initialized randomly, the model training takes a long time with CNN [10]. Transfer learning can increase the speed of training time by using the weights from the previous training activity architecture as an initial weight. The new architecture uses the new dataset and additional layers exactly like the previous architecture to transfer the weights. Those assigned weights can be maintained or changed during the training of the new architecture[11].

In the current research, we investigate the extension of Convolutional neural networks work (CNN) and Transfer Learning and how effective techniques are to extract hand features and interpret hand signs of PSL. We used the Visual Geometry Group (VGG) models with two unlike optimizers one is Stochastic Gradient Descent (SGD), and the second is Adam[12]. Considering VGG, we have selected the VGG16 model with a pre-trained architecture using the ImageNet data consisting of 1.2 million color images, and the number of classes is 1000[13]. By adapting the pre-trained model, we performed very well with minimal and low-quality data of 65 individuals having 455 images in total using Adam optimizer. Further, we analyzed the effect of VGG16 transfer learning architecture using Stochastic Gradient Descent (SGD) and Adam optimizers. Supplementary to another architecture AlexNet, ResNet-50, and LeNet-05 based on CNN were also tested and there is no successful model for our dataset but provided a competent architecture for hand sign classification. The current paper contributes in the following ways:

1. A CNN that interprets emotions through hand signs using Pakistan Sign Language
2. A publically available dataset for Pakistan Sign Language hand signs representing emotions
3. An approach based on transfer learning adapted the model with maximum accuracy with a minimal and low-quality dataset.
4. A comparison of AlexNet, LeNet05, ResNet50 and VGG16 architecture.
5. VGG16 architecture is evaluated with two different optimizers SGD and Adam.
6. A comparison of our model with existing state-of-the-art models.

Current section 01, based on the domain introduction and motivation of the present paper is composed of the structure as follows: Section 02 boons a background of Transfer Learning along with some deep learning architectures. Section 03 describes the methodology, where the dataset and adapted architecture are discussed. In Section 04, results and evaluations are being evaluated. Finally, Section 5 provides the conclusion and future directions.

BACKGROUND

In Sign Language, hand gesture and posture plays a vital role in understanding[14]. The recognition of hand gestures is the technique of analyzing a pattern in images. It generally requires four steps to recognize any motion, which are as follows: image acquisition: retrieves unprocessed labeled images from a source image pre-processing: performs grayscale conversion to several other techniques to represent images as matrices of pixels. Feature Extraction: in this step, analyzing image patterns are analyzed and extracted image classification: new or unseen images are checked or classified among predefined classes [15]. Several machine learning-based hand gesture recognition methods are in use [16]–[18]. However, an encouraging line of plans are in use nowadays are those techniques are based on Deep Learning and further built on CNNs [19]. Using CNN is particularly suitable for image recognition because it automates the feature extraction process and avoids poor generalization; it also decreases bias in traditional algorithms [20]. CNN takes a long time to converge by using random weights. The solution to this issue can use transfer learning for better accuracy and a prompt training time [21]. In the current study, we have used the VGG16 pre-trained model and proposed a model with a new dataset of hand gestures. The proposed architecture demonstrated the correct classification of all training and validation data. Table 01 describes the literature summary.

Table 1: Literature Summary

Year	Model / Approach	Sign Language	Signs	Accuracy	Description
2018 [20]	Restricted Boltzmann Machine	American Sign Language	Finger-spelling alphabets	99.31%	Model shows difficulty recognizing characters with low visual inter-class variability.

2019 [19]	Convolutional Neural Network	Bangla Sign Language	Finger-spelling numbers	92%	Low quality data can be improved by hand structuring devices and it will increase accuracy.
2020 [17]	Discrete Wavelet Transform (DWT) and Support Vector Machine (SVM)	American Sign Language	Random four signs	96.5%	Scope of the study and dataset is very low.
2021 [16]	Deep Convolutional Neural Network	American Sign Language	All English static alphabets	96.2%	A similar gesture usually occurs as a misclassification.
2022 [18]	Deep Convolutional Neural Network	American Sign Language	Finger-spelling for numbers	91.41%	An old dataset of numbers was used. The study has limitations in the pose and variance also.

Existing CNN architectures

Convolutional Neural Network (CNN) is renowned for its robust feature extraction and classification competencies. Introduced various architectures with a constant baseline learning mechanism for object recognition and profound learning mechanism improvements [22]. Among several famous architectures are AlexNet, LeNet-05, ResNet-50, and VGG16. Alex et al. proposed the AlexNet architecture, which is the first object recognition model that tries to learn network parameters over large-scale databases. There are 26 layers of AlexNet architecture; this architecture can easily be categorized into three sections. Figure 01 shows the AlexNet architecture, the first part consists of 2 units, and both comprise convolution, Relu (activation function), normalization, and pooling layers. The second part of the architecture includes four units consisting of 2 layers each; convolutional and pooling. The non-Linear activation unit was the base of the last section of architecture. Further corresponding to Fully Connected (FC) layers, Relu, and drop-out layers. A Drop-out layer is used to avoid overfitting during training, and the last two dealings of AlexNet are Softmax and Output for seven classes [23].

The accuracy of the CNN architecture, for the most part, depends on a vast dataset, high-end computational systems, and the depth of the network [24]. The last parameter is uncertain for AlexNet architecture due to the unavailability of such measures that could limit the network's depth.

Another convolutional architecture, LeNet-05 [25] is proposed by LeCun, initially for handwritten character recognition. This CNN-based model is capable enough of taking multiple objects as input and producing multiple outputs in a single pass without prior segmentation, which is called Space Displacement Neural Network (SDCNN) [26]. Sparse Convolutional layers and max-pooling layers are the souls of LeNet architecture. Details of the model are depicted in figure 02. From the implementation point of view, lower layers of the architecture work on 4D tensors, and a further flattened layer convert feature maps into the 2D matrix.

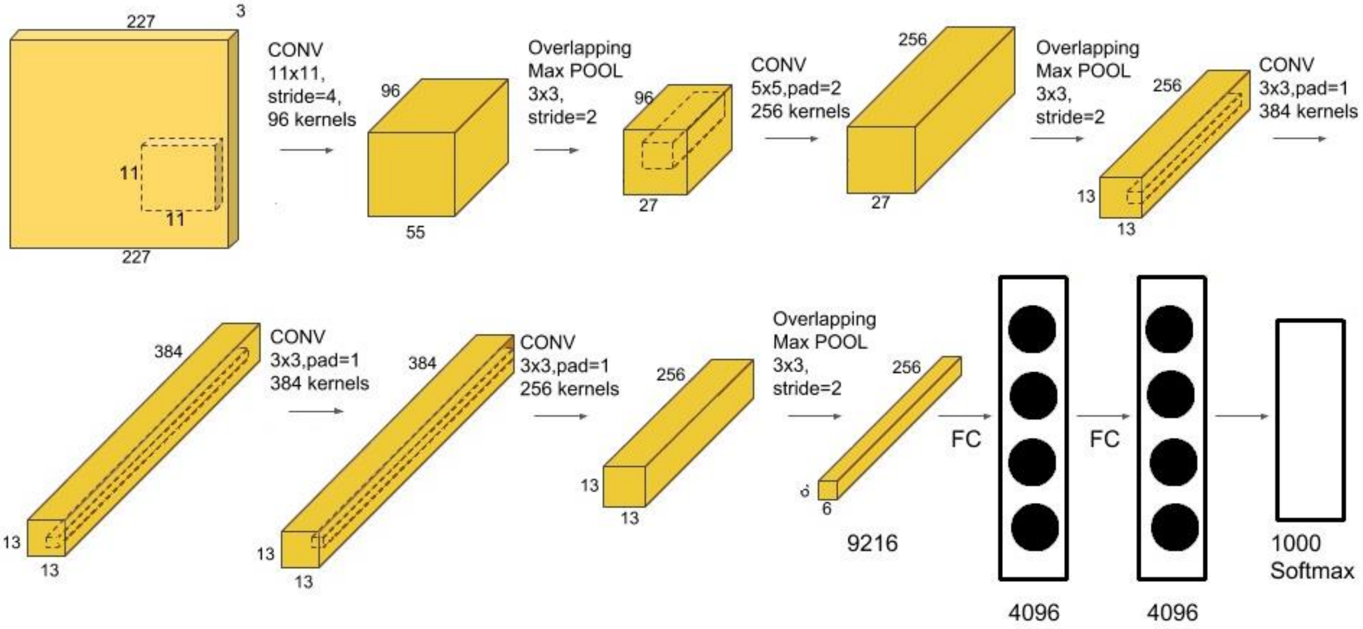


Figure 1: AlexNet Architecture

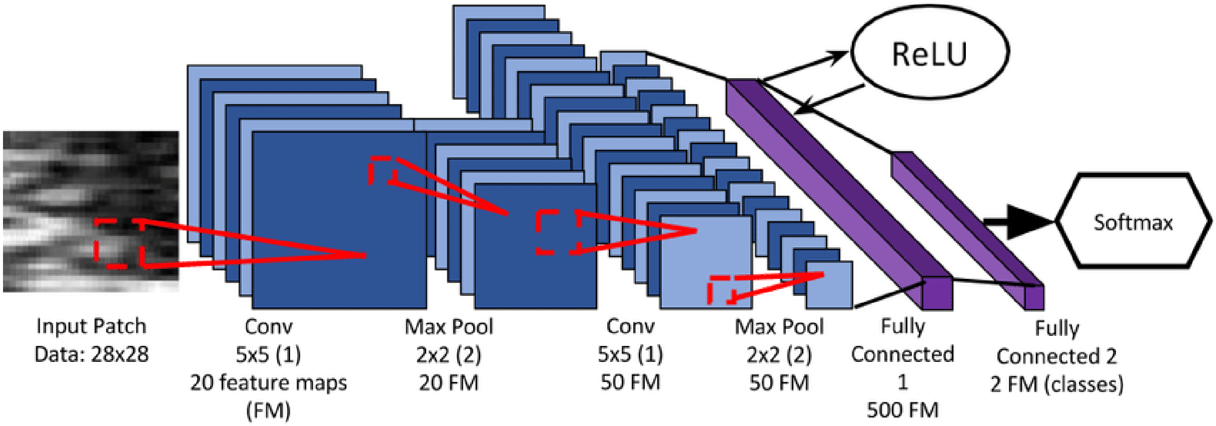


Figure 2: LeNet-5 Architecture

An alternative powerful adherent of the CNN family is ResNet50 architecture [27]. This model consists of 48 convolutional layers along with max pooling and average pooling layers one each. Figure 03 shows the model details. The essence of this model is based on a deep residual learning framework; it is specifically intended to solve vanishing gradient problems with extreme deep neural nets. This architecture has eminence due to many reasons but regardless of 50 layers ensuring around 23 million trainable parameters are making it a much smaller network than other existing CNN architectures.

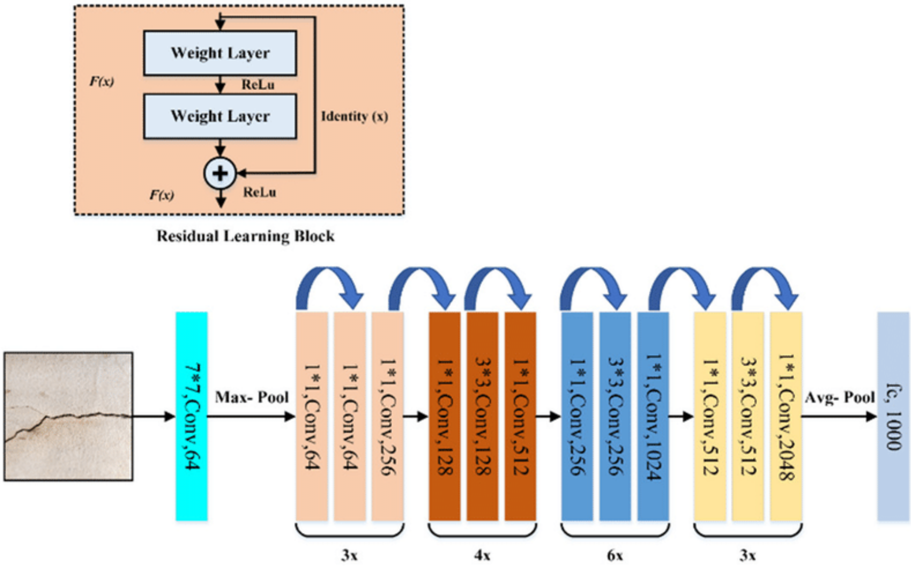


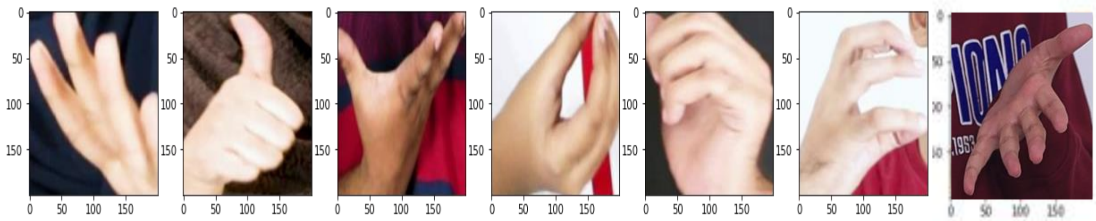
Figure 3: ResNet50 architecture

Here in the current study, we have proposed the architecture of VGG16 with a base CNN model for object detection that is described in the next section. In the architecture, unlike other networks, the Convolutional, Relu, and Pooling are replicative in the VGG16 deeper network. Also, they considered the smaller size of the receptive window of each convolutional layer [28].

METHODOLOGY

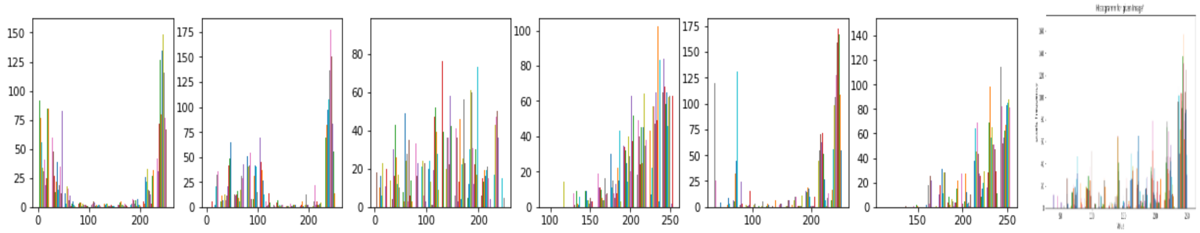
Hand gestures dataset

The collected dataset used in this research with a mobile device OPPO A76 having dual Cameras: 13MP, f/2.2 and LED flash. The selection of the gestures is based on the basic seven expressions of sad, happy, neutral, disgust, scared, angry, and surprise. Here to elaborate on the seven basic expressions through hand we have selected seven adjectives from Pakistan Sign Language, which are expressed as follows: disgust feeling is expressed by a bad adjective, the neutral feeling is expressed by the best adjective, the happy feeling is elaborated by glad as an adjective, the sad feeling is associated to the sad adjective, just like scared expression is associated with the scared adjective, the further the stiff adjective expresses the further angry feeling surprise expression has the same surprise adjective in PSL to portray action. Figure 04(a) depicts the seven basic expressions by hands defined in Pakistan Sign Language.



a. Bad b. Best c. Glad d. Sad e. Scarred f. Stiff g. Surprise

Figure 4(a): Seven hand gestures elaborating seven basic expressions using hands in PSL



a. Bad b. Best c. Glad d. Sad e. Scarred f. Stiff g. Surprise

Figure 4(b): Histograms of seven hand gestures elaborating seven basic expressions in PSL

All male and female individuals are from 20 to 50 years old. Each participant was asked to keep a frontal view to keep the hand posture natural and according to their body postures of each participant are reviewed and ensured the subject's performance and required suitable quality on the spot. Attire (shirt) or any other background is not controlled and has minor fluctuations in subjects' position and orientation. All other aspects of data collection were controlled. In the overall data set 48% of individuals used the right hand for gesture representation and the remaining 52% used the left hand. So, in this state, we can say that data is balanced and flipping is done, right or left-hand flipping causes any misclassification after model training and system completion.

As a preprocessing stage, all images after the acquisition were resized to 312x306 dimension Figure 04. (b) depicts the histogram representation of all hand gestures expressing a specific emotion. Further, the dataset splits in a 75:25 ratio for test and train data.

Convolutional Neural Network (CNN)

We cannot overlook Neural Nets when it comes to image recognition or vision-based machine learning propagation method is used to perform CNN training in certain specific layers. The convolutional layer is one of those hidden layers in the model [29]. The function of overall layers is feature extraction and later classifying those features at the end as res is treated as a tensor or high order matrix while moving inside the existing layers, on the other hand at certain layers training weights are stored which are called parameters. CNN has the following general layers [30]:

Input Layer

The input layer takes images as input directly. It is the high-order matrix storing dimensions of the image, those dimensions are length, width, number of channels, and transformations of an input image.

Convolutional Layer

The convolutional layer is mainly possible to apply the convolutional process, as well this layer stores the weights of training results. This layer provides a feature map with a smaller length and width but with more depth. Training weights as parameters can be calculated using the below:

$$N_{param} = (k1*k2*N_{input}*N_{output}+N_{bias}) \qquad \text{---------} \ (1)$$

In the above equation, k1 and k2 are kernel sizes, N_{input} and N_{output} are the numbers of input and output filters respectively, and N_{bias} is the number of biases which is the same as the number of N_{output} filters.

Activation Layer

This layer is the activation function for the convolutional layer. Rectified Linear Unit (Relu) is the usual function used due to its light nature, as it changes only negative input values to zero while positive input remains.

Pooling Layer

This layer retrieves the optimal features of the input tensor by subsampling from the previous layer. The output size of the kernel size determines the output size of the tensor. This layer works as a regularizer for the model. After downsizing the new matrix is a pooled feat, and a pooled feature map is the next layer.

Fully Connected Layers

This is the last layer of the model that performs as a classifier. When the overall model learns all the features, the final step is flattening the final pooled features by converting a feature map matrix into a single column matrix. Artificial neurons in this layer are trained and store the training results weights. The softmax activation function is used to decide the most dominant class, the formula to formulate the parameters as the convolutional layer with the kernel size 1 is given as equation (2)

$$N_{param} = (N_{input} * N_{output} + N_{bias}) \text{-----------} (2)$$

Figure 05. General architecture of CNN

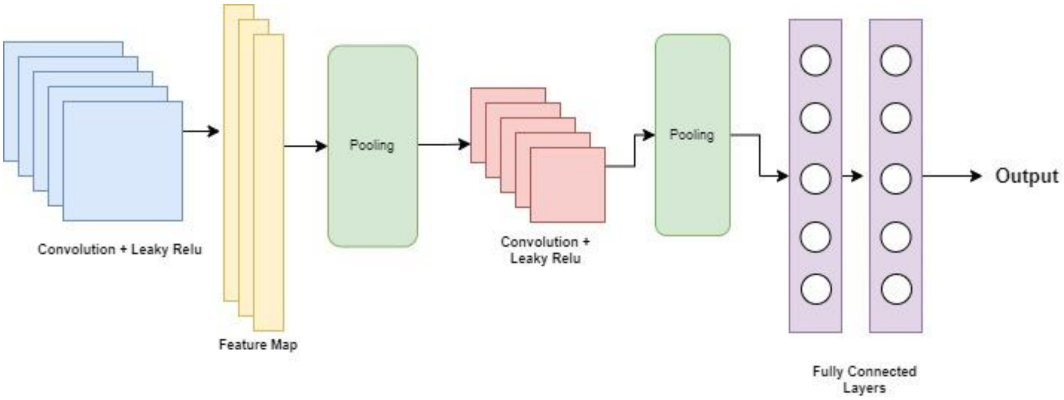


Figure 5 describes the general architecture of the Convolutional Neural Network (CNN). Here input image splits into many smaller matrices and is provided as input to the convolutional layer, after convolving and Relu function it performs the pooling process. These two processes are repeated several times to make the network dense. Finally, after feature learning the classification, the section consists of flattening and fully connected layers to provide output as classes in the final layers.

VGG-16 architecture using Transfer Learning

In this paper, we used the VGG16 model architecture. The original model is pre-trained with 1.2 million images having RGB color schemes for 1000 classes [31]. The basic model of VGG16 has 16 convolutional layers using 3x3 kernel sizes along with the Relu activation function. In architecture, each convolutional layer is followed by a max-pooling layer and the kernel size is always 2x2. The function of convolutional layers is to automate the task of feature extractions and storage of training weights. Further, the final layers as classifiers are 3 fully connected layers (FC). The number of parameters is determined collectively by convolutional layers and fully connected layers. Figure 06 defines the number of parameter output layers' number of parameters and tensor size generates the first 19 layers are extracting features, and layers 20 to 23 are performing the task of classification. Originally, the total number of parameters of VGG16 are very high at 134,289,223.

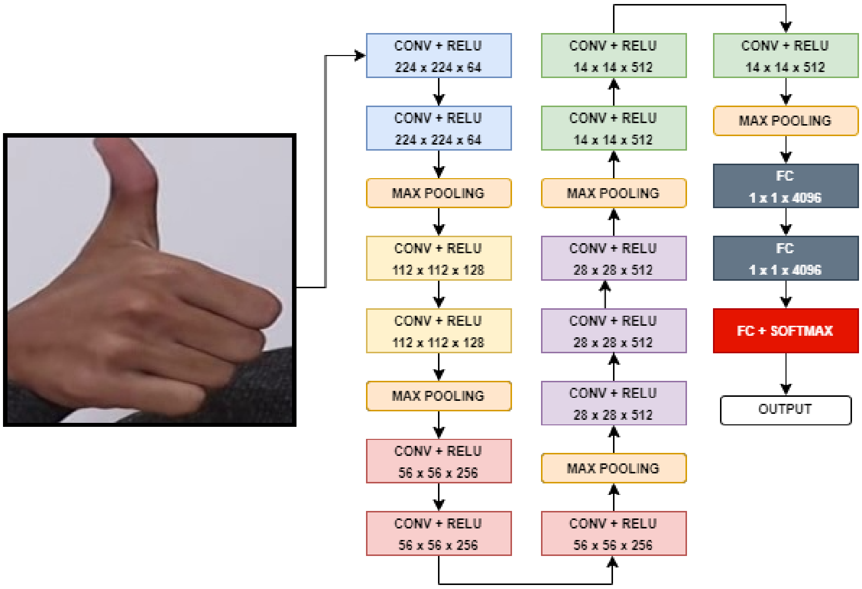


Figure 6. VGG16 transfer learning model

Proposed Architecture

The current study is based on a pre-trained VGG16 model. Our model consists of two sections: 1. feature extractor, and 2. classifier. Here the first section is used to extract features from the new dataset based on 7 hand gestures representing symbolic expressions in sign language. The classification part of the model is substituted with a new FC-layer to adjust the seven classes of the new dataset. Transfer learning is carried out in the feature-extraction section, where weights are convergent. Further, in the classification section, another FC-Layer is added to handle 7 classes of our study. Usually, the learning rate in transfer learning is always very minimal epochs, as most of the parameters are required to be convergent. In this current study, we used 25 epochs with a learning rate of 0.0001. Figure 07 elaborates on the difference between the original VGG16 architecture and the new proposed architecture adjusting the PSL hand expression dataset.

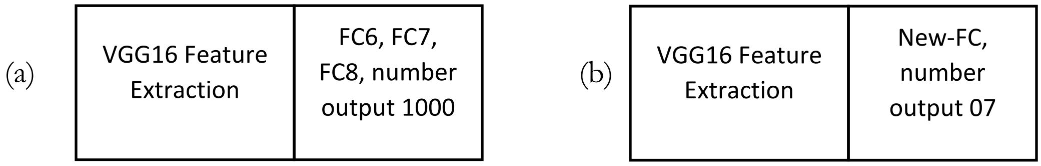


Figure 07: (a) original VGG16 model, (b) proposed model

RESULTS AND DISCUSSIONS

To analyze the test scenario for our finned tuned architecture of VGG16 with Adam optimizer, results are represented graphically to make them clearer. More specifically (a) and (b) in figure 08 show the accuracy and loss graphs, which follow the training graph very well and achieve higher accuracy with only 25 epochs. Besides the accuracy and loss matrices we have, we have also used class-wise Average, Precision, and F1-score, Area Under the Curve (AUC) along with confusion matrix.

Accuracy is the ratio between total observations and correct predictions. Our accuracy is near 100 percent which, shows that our model is best for our case, but still, for any irregularity, we must check other parameters to justify the performance of the model. Equation (3) to (6) represents the formulas for all evaluation matrices.

$$Accuracy = \frac{True\ Positive + True\ Nagative}{True\ Positive + False\ Positive + False\ Negative + True\ Negative} \quad (2)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (3)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (4)$$

$$F1\ Score = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \quad (5)$$

To measure the model as a classifier, we are using Area Under the Curve (AUC) which is the summary of the ROC curve. The higher the AUC curve, the higher the accuracy to qualify any class as positive or negative.

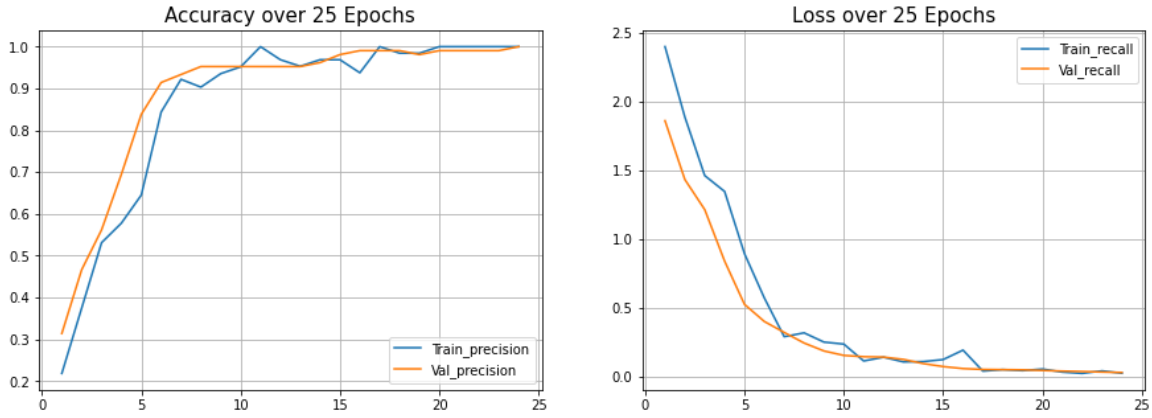
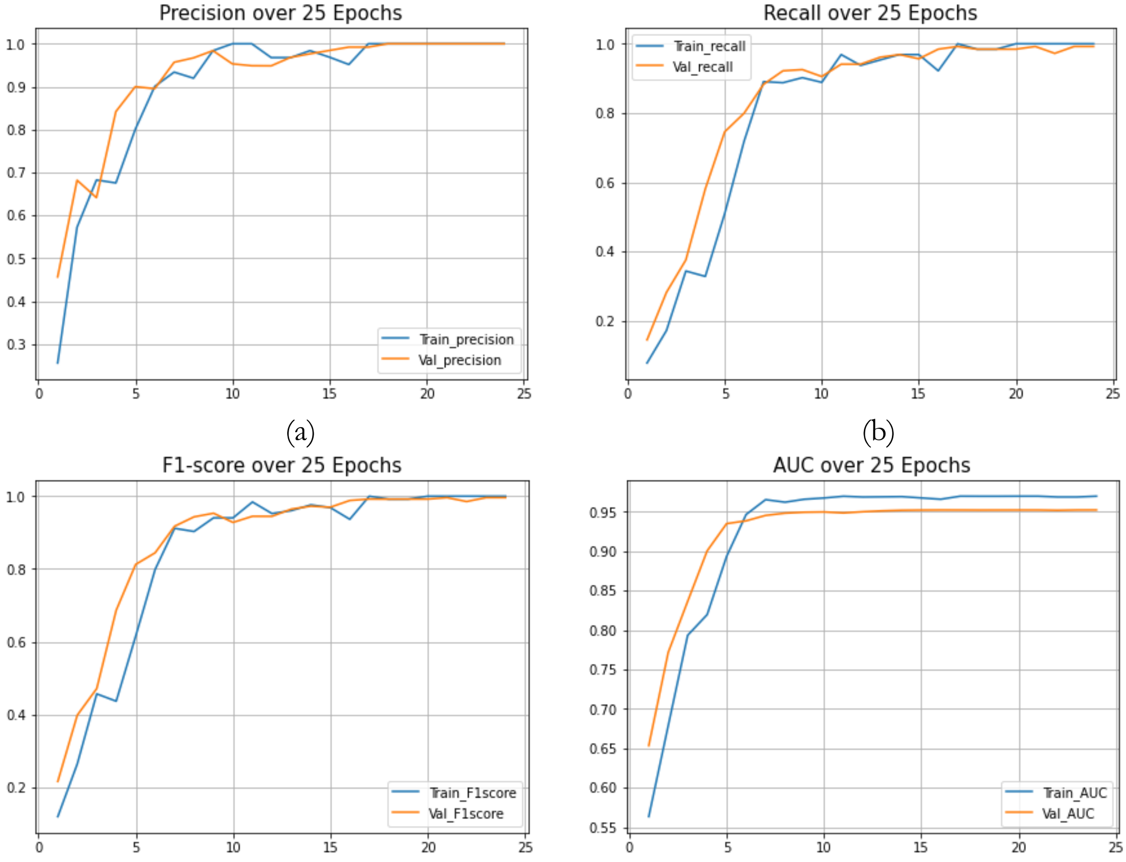


Figure 08 (a) (b)

(a) Accuracy graph for training and validation of the transfer learning model

(b) Loss graph for training and validation of the transfer learning model

Further, figure 09 represents Precision, Recall, F1 Score, and AUC. Here precision and recall are the ratios between total positive predictions, and overall observations and positive predictions respectively. The recall is also called the sensitivity of the model. Subsequently, the weighted average between two previously calculated ratios of precision and recall is the F1 score of the given model, which gives better results as compared to accuracy even with imbalanced classes. These all ratios are at the ideal condition of 1 using our tuned transfer learning model.



(a) (b)

(c) (d)

Figure 9 (a) Precision (b) Recall (c) F1-Score (d) AUC graphs

Further, in figure 09 (d) Area Under the Curve (AUC) shows 0.96 value for training and 0.95 value for validation representation. When AUC is between 0.5 to 1, it shows the classifier is able to distinguish the positive classes. For our case, results are far promising and prove the current model fit for the hand gesture classification problem. Similarly, in figure 10 Confusion matrix depict the error matrix and gives a remarkable result of 100 Percent accurate predictions.

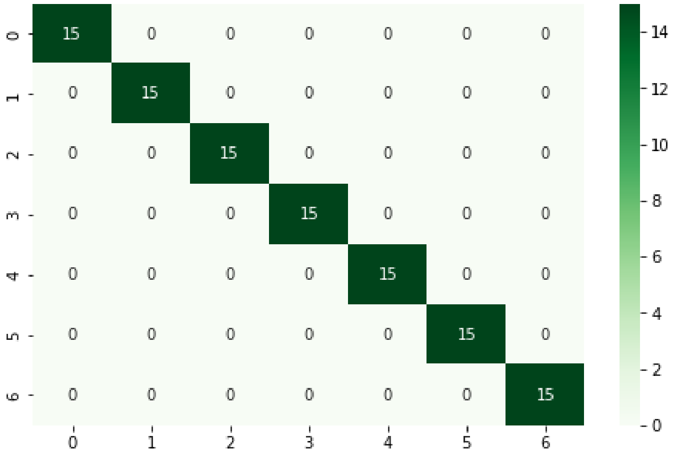


Figure 10: Confusion Matrix / Error Matrix

Moreover, in figure 10 prediction and testing of class-wise results are demonstrated to analyze the robustness of the model using transfer learning. All classes are classified between 99 to 100% valid class detection over these facts that hand orientations, backgrounds, brightness, and other features are variated for the overall dataset, our results are quite promising. Figure 10 predicts several hand gestures of PSL emotion-based hand gesture recognition accurately.

In the overall study, our analysis is based on three experiments. All three experiments are carried out using a dataset of Pakistan Sign Language, depicting human emotions based on hand gestures. Seven basic emotions of sad, bad, happy, neutral, disgust, anger, and surprise are considered. VGG16 architecture with pertained weights and Adam optimizer used for transfer learning on this image classification task. In the current and previous sections, this technique is elaborated well as it was evaluated that it is providing the best average results in terms of "Accuracy", "Loss", "Precision", "Recall", and," F1-Score". Further, in the second experimentation VGG16 architecture is used with all constraints of transfer learning but with an SGD optimizer. And in the third experimentation, three more state-of-the-art architecture of AlexNet, LeNet05, and ResNet50 is used for our image classification task. Both, the second and third experimentations are not so promising. Table 02 shows a comparison between CNN architectures using accuracy, precision, recall, and F1-Score as an average evaluation score for our seven class's image recognition tasks. Table 02 witnessed the accuracy of VGG16 architecture with Adam Optimizer among other architectures.

Table 2: comparison between AlexNet, VGG16 Adam optimizer, and VGG16 SGD optimizer

Method	Accuracy	Precision	Recall	F1-Score
AlexNet	96.88	96.56	96.45	96.55
LeNet05	89.45	88.76	88.54	89.01
ResNet50	97.32	97.55	97.58	97.56
VGG16 (SGD)	97.12	97.23	97.22	97.10
VGG16 (Adam)	**99.98**	**99.98**	**99.98**	**99.89**

The overall classification accuracy of our model is 100%, which is definitely at maximum than existing work in this current domain. While using more challenging and robust dataset in terms of small and low quality; Table 03 defines a comparison of model performances of our model with other publish state-of-the-artwork.

Table 03: Comparison of model performances with other published work

Ref. No.	Classification Technique	Classification Accuracy
[32]	Image Net + VGG16 trained network	94.56%
[33]	DNN + Transfer Learning	93.5%
[34]	CNN + Ensemble Model	96%
[3]	3DCNN+LSTM	93.09%
Our Model	CNN +VGG16 (with Adam Optimizer)	100%

We have selected the deep learning approach and proposed a network as the importance of the Artificial Intelligence, Machine Learning and Deep Learning cannot be neglect in this era. Machine Learning and Deep learning are being used in different area like Medical, Computer Vision, Sentiment, Natural Language Processing and Internet of Things [35]–[40].

CONCLUSION AND FUTURE DIRECTIONS

In the current study, we used transfer-learning based state-of-art-networks AlexNet, LeNet05, ResNet50, VGG16 with SGD optimizer, and VGG16 with Adam optimizer over a newly developed dataset for Pakistan Sign Language, consisting of seven classes of expression represented by hand gestures. The performance of these networks is measured by accuracy, loss, Precision, Recall, F1-Score, and AUC. Performance analysis witnessed improvement in VGG16 architecture while using Adam Optimizer. In the future, these fine-tuned network architectures can be used for real-time sign language recognition systems. Additionally, sign language combines of manual and non-manual gestures, manual gestures are gestures performed using hand signs either single or double. On the other hand, non-manual signs are gestures based on body movement, head movement, torso, and facial expressions. The current study besieged seven basic expressions of sad, happy, disgust, anger, neutral, scared, and surprise expressions through hands but it is very important to combine facial expressions and body movements along with such adjectives to enhance the meaning of such expressions using hand and facial features. So, in the future current architecture can be used for human action recognition and specifically for sign language action recognition, non-manual sign language recognition, or for the hybrid model for manual and non-manual hybrid sign language modalities recognition.

References

[1] R. Vaccaro, D. Zaccaria, M. Colombo, S. Abbondanza, and A. Guaita, "Adverse effect of self-reported hearing disability in elderly Italians: Results from the InveCe.Ab study," *Maturitas*, vol. 121, pp. 35–40, Mar. 2019, doi: 10.1016/J.MATURITAS.2018.12.009.

[2] R. Rastgoo, K. Kiani, and S. Escalera, "Real-time isolated hand sign language recognition using deep networks and SVD," *J. Ambient Intell. Humaniz. Comput. 2021 131*, vol. 13, no. 1, pp. 591–611, Feb. 2021, doi: 10.1007/S12652-021-02920-8.

[3] R. Rastgoo, K. Kiani, and S. Escalera, "Hand sign language recognition using multi-view hand skeleton," *Expert Syst. Appl.*, vol. 150, p. 113336, Jul. 2020, doi: 10.1016/J.ESWA.2020.113336.

[4] R. Rastgoo, K. Kiani, and S. Escalera, "Sign Language Recognition: A Deep Survey," *Expert Syst. Appl.*, vol. 164, p. 113794, Feb. 2021, doi: 10.1016/J.ESWA.2020.113794.

[5] H. Zahid, M. Rashid, S. Hussain, F. Azim, S. A. Syed, and A. Saad, "Recognition of Urdu sign language: a systematic review of the machine learning classification," *PeerJ. Comput. Sci.*, vol. 8, 2022, doi: 10.7717/PEERJ-CS.883.

[6] E. De Stefani and D. De Marco, "Language, gesture, and emotional communication: An embodied view of social interaction," *Front. Psychol.*, vol. 10, no. SEP, pp. 1–8, 2019, doi: 10.3389/fpsyg.2019.02063.

[7] T. R. Gadekallu *et al.*, "Hand gesture classification using a novel CNN-crow search algorithm," *Complex Intell. Syst.*, vol. 7, no. 4, pp. 1855–1868, 2021, doi: 10.1007/s40747-021-00324-x.

[8] S. Javaid, S. Rizvi, M. T. Ubaid, and A. Tariq, "VVC/H.266 Intra Mode QTMT Based CU Partition Using CNN," *IEEE Access*, vol. 10, pp. 37246–37256, 2022, doi: 10.1109/ACCESS.2022.3164421.

[9] C. M. Sharma, K. Tomar, R. K. Mishra, and V. M. Chariar, "Indian Sign Language Recognition Using Fine-tuned Deep Transfer Learning Model," *SSRN Electron. J.*, Sep. 2021, doi: 10.2139/SSRN.3932929.

[10] S. J. Goyal, A. K. Upadhyay, and R. S. Jadon, "Combined Approach to Classify Human Emotions Based on the Hand Gesture," pp. 301–309, 2020, doi: 10.1007/978-981-15-0633-8_29.

[11] Y. Zou and L. Cheng, "A Transfer Learning Model for Gesture Recognition Based on the Deep Features Extracted by CNN," *IEEE Trans. Artif. Intell.*, vol. 2, no. 5, pp. 447–458, Jul. 2021, doi: 10.1109/TAI.2021.3098253.

[12] A. Ranjan, C. Kumar, R. K. Gupta, and R. Misra, "Transfer Learning Based Approach for Pneumonia Detection Using Customized VGG16 Deep Learning Model," *Lect. Notes Networks Syst.*, vol. 340 LNNS, pp. 17–28, 2022, doi: 10.1007/978-3-030-94507-7_2/COVER/.

[13] M. Shaha and M. Pawar, "Transfer Learning for Image Classification," *Proc. 2nd Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2018*, pp. 656–660, Sep. 2018, doi: 10.1109/ICECA.2018.8474802.

[14] F. Noroozi, C. A. Corneanu, D. Kaminska, T. Sapinski, S. Escalera, and G. Anbarjafari, "Survey on Emotional Body Gesture Recognition," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 505–523, Jan. 2018, doi: 10.48550/arxiv.1801.07481.

[15] M. A. Arjun, S. Sreehari, and R. Nandakumar, "The Interplay of Hand Gestures and Facial Expressions in Conveying Emotions A CNN-BASED APPROACH," *Proc. 4th Int. Conf. Comput. Methodol. Commun. ICCMC 2020*, pp. 833–837, Mar. 2020, doi: 10.1109/ICCMC48092.2020.ICCMC-000154.

[16] R. Bhadra and S. Kar, "Sign language detection from hand gesture images using deep multi-layered convolution neural network," *2021 IEEE 2nd Int. Conf. Control. Meas. Instrumentation, C. 2021 - Proc.*, pp. 196–200, Jan. 2021, doi: 10.1109/CMI50323.2021.9362897.

[17] P. Parvathy, K. Subramaniam, G. K. D. Prasanna Venkatesan, P. Karthikaikumar, J. Varghese, and T. Jayasankar, "Development of hand gesture recognition system using machine learning," *J. Ambient Intell. Humaniz. Comput. 2020 126*, vol. 12, no. 6, pp. 6793–6800, Jul. 2020, doi: 10.1007/S12652-020-02314-2.

[18] S. Bhushan, M. Alshehri, I. Keshta, A. K. Chakraverti, J. Rajpurohit, and A. Abugabah, "An Experimental Analysis of Various Machine Learning Algorithms for Hand Gesture Recognition," *Electron. 2022, Vol. 11, Page 968*, vol. 11, no. 6, p. 968, Mar. 2022, doi: 10.3390/ELECTRONICS11060968.

[19] S. Ahmed *et al.*, "Hand Sign to Bangla Speech: A Deep Learning in Vision Based System for Recognizing Hand Sign Digits and Generating Bangla Speech," *SSRN Electron. J.*, pp. 1–6, 2019, doi: 10.2139/ssrn.3358187.

[20] R. Rastgoo, K. Kiani, and S. Escalera, "Multi-Modal Deep Hand Sign Language Recognition in Still Images Using Restricted Boltzmann Machine," *Entropy 2018, Vol. 20, Page 809*, vol. 20, no. 11, p. 809, Oct. 2018, doi: 10.3390/E20110809.

[21] J. Pardede, B. Sitohang, S. Akbar, and M. L. Khodra, "Implementation of Transfer Learning Using VGG16 on Fruit Ripeness Detection," *Int. J. Intell. Syst. Appl.*, vol. 13, no. 2, pp. 52–61, 2021, doi: 10.5815/ijisa.2021.02.04.

[22] S. Arora, A. Gupta, R. Jain, and A. Nayyar, "Optimization of the CNN Model for Hand Sign Language Recognition Using Adam Optimization Technique," *Lect. Notes Networks Syst.*, vol. 179 LNNS, pp. 89–104, 2021, doi: 10.1007/978-981-33-4687-1_10/COVER/.

[23] L. and W. Ministry of Health, "The History Began from AlexNet: A Comprehensive

Survey on Deep Learning Approaches," 2012, [Online]. Available: http://www.mhlw.go.jp/new-info/kobetu/roudou/gyousei/anzen/dl/101004-3.pdf.

[24] A. Jain, A. Sethi, D. K. Vishwakarma, and A. Jain, "Ensembled Neural Network for Static Hand Gesture Recognition," *2021 12th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2021*, 2021, doi: 10.1109/ICCCNT51525.2021.9579633.

[25] S. Mascarenhas and M. Agarwal, "A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification," *Proc. IEEE Int. Conf. Disruptive Technol. Multi-Disciplinary Res. Appl. CENTCON 2021*, pp. 96–99, 2021, doi: 10.1109/CENTCON52345.2021.9687944.

[26] M. L. George, T. Govindarajan, K. Angamuthu Rajasekaran, and S. R. Bandi, "A robust similarity based deep siamese convolutional neural network for gait recognition across views," *Comput. Intell.*, vol. 36, no. 3, pp. 1290–1319, Aug. 2020, doi: 10.1111/COIN.12361.

[27] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.

[28] A. Chavan, J. Ghorpade-Aher, A. Bhat, A. Raj, and S. Mishra, "Interpretation of Hand Spelled Banking Helpdesk Terms for Deaf and Dumb Using Deep Learning," *2021 IEEE Pune Sect. Int. Conf. PuneCon 2021*, 2021, doi: 10.1109/PUNECON52575.2021.9686514.

[29] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 24–49, Mar. 2021, doi: 10.1016/J.ISPRSJPRS.2020.12.010.

[30] S. Ghaffarian, J. Valente, M. Van Der Voort, and B. Tekinerdogan, "Effect of Attention Mechanism in Deep Learning-Based Remote Sensing Image Processing: A Systematic Literature Review," *Remote Sens. 2021, Vol. 13, Page 2965*, vol. 13, no. 15, p. 2965, Jul. 2021, doi: 10.3390/RS13152965.

[31] N. Nguyen Tu, S. Sako, and B. Kwolek, "Fingerspelling recognition using synthetic images and deep transfer learning," no. March, p. 70, 2021, doi: 10.1117/12.2587592.

[32] A. P. Parameshwaran, H. P. Desai, R. Sunderraman, and M. Weeks, "Transfer learning for classifying single hand gestures on comprehensive bharatanatyam mudra dataset," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2019-June, pp. 508–510, Jun. 2019, doi: 10.1109/CVPRW.2019.00074.

[33] K. Suri and R. Gupta, "Transfer learning for sEMG-based hand gesture classification using deep learning in a master- slave architecture," *Proc. 3rd Int. Conf. Contemp. Comput. Informatics, IC3I 2018*, pp. 178–183, Oct. 2018, doi: 10.1109/IC3I44769.2018.9007304.

[34] A. P. Parameshwaran, H. P. Desai, M. Weeks, and R. Sunderraman, "Unravelling of Convolutional Neural Networks through Bharatanatyam Mudra Classification with Limited Data," *2020 10th Annu. Comput. Commun. Work. Conf. CCWC 2020*, pp. 342–347, Jan. 2020, doi: 10.1109/CCWC47524.2020.9031185.

[35] M. A. Arshed, H. Ghassan, M. Hussain, M. Hassan, A. Kanwal, and R. Fayyaz, "A Light Weight Deep Learning Model for Real World Plant Identification," *2022 2nd Int. Conf. Distrib. Comput. High Perform. Comput. DCHPC 2022*, pp. 40–45, 2022, doi: 10.1109/DCHPC55044.2022.9731841.

[36] M. T. Ubaid, M. Z. Khan, M. Rumaan, M. A. Arshed, M. U. G. Khan, and A. Darboe,

"COVID-19 SOP's Violations Detection in Terms of Face Mask Using Deep Learning," *4th Int. Conf. Innov. Comput. ICIC 2021*, 2021, doi: 10.1109/ICIC53490.2021.9692999.

[37] B. Liu *et al.*, "Comparison of Machine Learning Classifiers for Breast Cancer Diagnosis Based on Feature Selection," *Proc. - 2018 IEEE Int. Conf. Syst. Man, Cybern. SMC 2018*, pp. 4399–4404, Jan. 2019, doi: 10.1109/SMC.2018.00743.

[38] M. T. Ubaid, A. Kiran, M. T. Raja, U. A. Asim, A. Darboe, and M. A. Arshed, "Automatic Helmet Detection using EfficientDet," *4th Int. Conf. Innov. Comput. ICIC 2021*, 2021, doi: 10.1109/ICIC53490.2021.9693093.

[39] O. C. Uner, H. I. Kuru, R. G. Cinbis, O. Tastan, and E. Cicek, "DeepSide: A Deep Learning Approach for Drug Side Effect Prediction," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, no. 2016, 2022, doi: 10.1109/TCBB.2022.3141103.

[40] D. Chaudhary, D. Karim, H. Alam, and S. Mumtaz, "Machine Learning with Data Balancing Technique for IoT Attack and Anomalies Detection Original Article," *Int. J. Innov. Sci. Technol.*, vol. 4, no. 2, pp. 490–498, 2022.