IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

SignSense: AI Framework for Sign Language Recognition

Prof. V. M. Dilpak[1], Rewa S. Joshi[2], Harshada K. Sonje[3]
Professor, Department of AI & ML[1]
Students, Department of AI & ML[2,3]
AISSMS Polytechnic, Pune, India

Abstract: *Sign Language recognition is a pioneering framework designed to advance the field of Sign Language Recognition (SLR) through the innovative application of ensemble deep learning models. The primary goal of this research is to significantly improve the accuracy, resilience and interpretability of SLR systems. Leveraging the unique features of ResNet within an ensemble learning paradigm. The key component of InceptionResNetv2 architecture is its deep and effective feature extraction capabilities. The utilization of InceptionResNet model enhances the model ability to capture intricate details crucial for accurate sign language recognition. This framework is also to scale seamlessly, accommodating an expanding vocabulary of signs, diverse users and dynamic environmental conditions without compromising performance.*

Keywords: Deep learning, computer vision, explainable AI, Sign Explainer, classification, sign language, technological development

I. INTRODUCTION

1.1 Existing System

Sign language plays a crucial role in fostering communication accessibility for the deaf and hard of hearing community. Deep learning techniques have significantly outperformed more the typical machine learning approaches, in various fields like computer vision, Natural language processing. Several CNN based DL model has developed for sign language recognition. The main challenges posed by existing Sign recognition methods is robust and accurate detection. To enhancing the recognition capability of SLR system, the existing system makes use of attention-based ensemble learning. The potential disadvantages associated with self-attention model is computational intensity, model size and storage.

1.2 Proposed System:

Sign language recognition is a critical aspects of facilitating communication for the deaf and hard-of- hearing community. This research introduces an innovative approach utilizing the inceptionResNetV2 architecture for robust and accurate sign language classification. A diverse dataset is employed, encompassing a variety of sign language gestures captured from real world scenarios. The proposed methodology involves the transfer learning technique, adapting the InceptionResNetV2 model to the specific requirements of sign language recognition. The model is fine-tuned with a new classification layer to suit the unique characteristic of the dataset. Additionally, data augmentation techniques are applied to enhance the models generalization capability. Training and Validation are conducted on a carefully partitioned dataset and the performance is evaluated through a comprehensive set of metrics. The resulting model demonstrates high accuracy, showcasing its potential for practical applications in real time sign language recognition scenarios. The study contributes to the ongoing efforts in developing efficient human-machine communication system for the deaf and hard of hearing community. The InceptionResNetV2 based model offers promising results, emphasizing its efficiency in advancing the state of the art in sign language recognition technology.

DOI: 10.48175/IJARSCT-17257

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

Modules:

- Dataset collection:
- DL model development and configuration

Dataset:

The sign language recognition system was assessed using ensemble learning on the Indian Sign Language (ISL) Dataset (kaggle.com) which contains 36 classes, including digits (0-9)and the alphabet
(A-Z). This dataset comprises around 1200 images for each class, with images having three channels.

InceptionResnetV2:

InceptionResNetV2 is a deep convolutional neural network architecture that combines the concepts of the inception module and residual connections. It was designed to enhance the feature learning capabilities while addressing challenges like vanishing gradients in very deep networks. The model is known for its impressive performance in image classification and object detection tasks. In transfer learning, InceptionResNetV2 is often employed as a pre-trained model on large dataset such as ImageNet. The pre-training on a diverse dataset allows the model to learn rich hierarchical features and representations. For a specific task, like sign language recognition in this context, the pre-trained InceptionResNetV2 model is fine-tuned on a smaller dataset related to the target domain. This fine-tuning process adjusts the model's parameters to adapt its knowledge to the nuances and characteristics of the new dataset. During transfer learning, the original classification layer of InceptionResNetV2 is typically replaced with a new one suited for the specific task at hand. The modified model is then trained on the target dataset, often with additional techniques like data augmentation to enhance its generalization ability. This transfer learning approach leveraging InceptionResNetV2 enables efficient and effective model training for specialized tasks with limited data, making it a popular choice in various computer vision applications.

Software Requirements:

- Operating System: Windows 10 (64 bit)
- Software: MATLAB

Hardware Requirements:

- Hard Disk : 500GB and Above
- RAM: 4GB and Above
- Processor: I3 and Above

Sign language recognition using transfer learning with InceptionResNetV2 architecture has presented in this work. The steps involved in SLR is dataset preparation, visualization, model configuration and training evaluation. Here are the main finding and potential points to include in this work

1. Model Architecture: The InceptionResNet V2 model is employed as a pre-trained feature extractor, harnessing its capability to capture intricate features from large datasets. The researchers modified the model by replacing the classification layer to align it with the specific task of sign language recognition.
2. Data Augmentation: To enhance model generalization, data augmentation techniques are applied

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

Architecture Diagram:

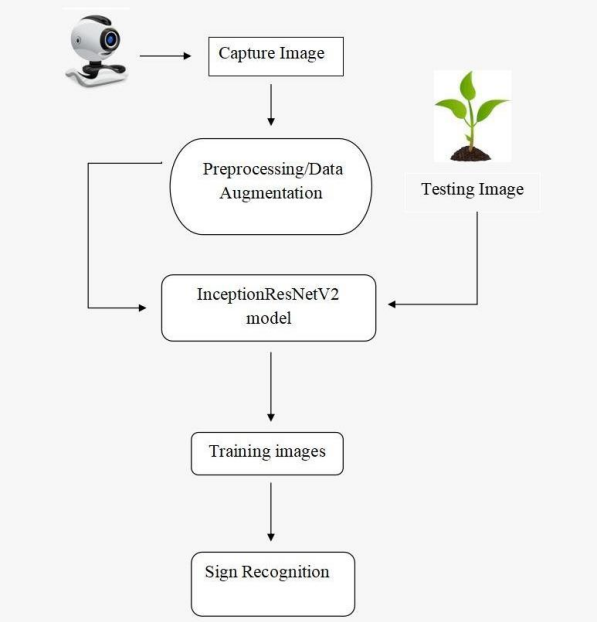

FIGURE 1. Architecture Diagram.

II. RELATED WORK

Kim et al. [11], introduce Concept Activation Vectors (CAVs), which translate a neural network's internal state into understandable ideas, which the author introduces. The important concept is to use a neural network's high-dimensional internal state as a tool rather than a hindrance. The authors have demonstrated the application of CAVs as a component of a method called Testing with CAVs (TCAV), which uses directional derivatives to gauge. How important a user-defined concept is to the categorization result, such as how much of a zebra prediction is influenced by the presence of stripes. We explain how CAVs may be used to evaluate predictions and generate knowledge for a standard image classification network and a medical application, putting con- cepts to the test in image categorization.

In this research [12], authors describe a unique technique that offers contrasting justifications for categorizing an input by a deep neural network or another black box classifier. Given an input, we find what needs to be simply and adequately present (viz. important object pixels in an image) to justify its classification and analogously, along with that minimally and necessarily absent (viz. certain background pixels) for the same. We contend that such explanations are typical in fields like criminology and health care because they are natural to people. A key aspect of an explanation that, to our knowledge, has not yet been formally identified by current explanation methods used to explain neural net- work predictions is minimally represented but critically not present. The authors have validated the proposed methodology over three datasets obtained from diverse domains; a brain activity strength dataset, a large procurement fraud dataset, and a handwritten digits dataset MNIST. In all three cases, we observe the effectiveness of our method in producing precise explanations that are also simple for specialists to comprehend and evaluate. [12].

Akula et al. [13], proposed the CoCoX model to explain the prediction generated by CNN classification. The author has proposed a fault-line model to identify minimum segmented- level features. Explanation from the CoCoX model was understandable to the technical and non- technical communities. The author has evaluated qualitative matrices like Justification Trust (JT), and Explanation Satisfaction (ES) to make performance understandable. The author has also compared the fault line model to other state-of-the-art models like LIME and LRP [13], author has successfully achieved 69.1 JT with CNN learning and Fault-Line Identification.

Contreras et al. [14], design Deep Explainer and Rule Extraction (DEXiRE), to make binary neural networks explainable. The proposed methodology uses rule extraction, which improves knowledge extraction from DL model (CNN) output. A final (global) rule set describing the general behav- ior of DL predictors can be created by integrating

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

intermediate rule sets explaining the behavior of each concealed layer. They used BCWD, Banknote, and Prima diabetes datasets for the simulation of the proposed DEXiRE model. The number of words in the intermediate and final rule sets may be regulated precisely with DEXiRE. The rule Extraction model has achieved remarkable accuracy and fidelity 0.94 and 0.95 respectively in a very small amount of time (around 232 ms).

Patel et al. [15] water Potability prediction synthetic over- sampling technique and Explainable AI. The author has used Synthetic Minority Oversampling Technique (SMOTE) method to classify water quality on the Kaggle dataset. The author has also compared the proposed architecture with other standard machine learning models

like Design Tree, Gradient Boost, Support vector machine, Random Forest, and Ada Boost. The proposed methodology has achieved 81% remarkable accuracy. The author has also considered the lack of transparency issue for Machine Learning models. To determine the significance of the characteristics of the predicted result, Local Interpretable Model-agnostic Expla- nations (LIME) are used. The author has demonstrated the different available particles in water like Chloramines, Tur- bidity, Sulfate, and many more to justify results with Explain- able AI, the proposed LIME model utilize to generate a result with the percentage of water particles.

Vermeire et al. [16] proposed a model-agnostic model ''Search for EviDence Counterfactual'' (SEDC) for image classification. The ''EdC'' explanation is an irreducible col- lection of characteristics that, if absent, would change the classification of the document. The SEDC additionally sup- ports a single task for image explanation. The proposed methodology used image segmentation as a core component to interpret. The authors have the simulated model to compare different counterfactual classes and also compare with standard explainer models like SHAP and LIME. Simulation has used pre-train weights of MobileNet V2 to demonstrate the interpretation of the proposed SEDC model.

Goel et al. [17], a proposed technique to design ''coun- terfactual explanations''. Generally, it is used to justify by content area of the image, through the model that made the prediction. The methodology also encountered the problem of Minimum-Edit Counterfactual. A methodology work on input image trained by a computer vision model, to interpret the predicted class. The methodology used the MNIST dataset over the CNN model achieved 98.40% accuracy. The proposed training model has 2 convolutions and 2 FC (Fully connected) layers to generate a feature size of $4 \times 4 \times 40$. To generalize counterfactual explanations, the author has also experimented with Omniglot and Caltech-UCSD Birds dataset. Proposed technique working over Greedy Sequential Exhaustive Search model. The author has summarized the qualitative and quantitative results of the proposed technique.

TABLE 1. Comparative analysis of state-of-the-art Explainable AI model overconfidence and justified trust value.

Author	Model	Justified Trust	Confidence
Zhou et al. 2016 [19]	CAM	37.1% ± 3.9%	3.2 ± 1.8
Selvaraju et al. 2017 [20]	Grad-CAM	39.1% ± 2.1%	3.7 ± 1.2
Ribeiro, Singh, and Guestrin 2016 [21]	LIME	42.1% ± 3.1%	3.1 ± 2.2
Kim et al. 2018 [11]	TCAV	55.1% ± 3.3%	3.9 ± 2.8
Dhurandhar et al. 2018 [12]	CEM	61.1% ± 2.2%	4.8 ± 1.6
Goyal et al. 2019 [17]	CVE	64.5% ± 3.7%	4.1 ± 2.3
Akula et al. 2020 [13]	CoCoX	70.5% ± 1.3%	5.7 ± 1.1
Vermeire et al. 2022 [16]	SEDC	71.4% ± 2.1%	6.1± 1.0

Arras et al. [18], proposed a framework that provides, a controlled, selective, and realistic testbed for the prediction of deep neural networks. The proposed methodology uses the CLEVR-XAI dataset for simulation, there were around 140k questions in the CLEVR-XAI evaluation set. With 28 alternative solutions. The prediction issue is presented as a classification challenge. The author has used ten polling techniques to visualize the explanation evaluation over a round truth mask. The experiment section summarized the evaluation of different XAI methods like Guided Backprop, LRP, SmoothGrad, and other 7 methods [18]. The conclusive study finds that LRP performed much better compared to another method over the proposed (CLEVR-XAI) benchmark dataset. Table 1 represent comparative analysis over

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

different explainable model to predict result by black-box learning, analysis also represents a statistical comparison to justify trust and confidence

III. MATERIALS AND METHODS

The proposed architecture used an Explainable AI-based methodology for sign language recognition with DeepExplainer. Which use to predict and validate generated output with learning interpretability. The proposed methodology uses SHAP (Shapley Additive exPlanations) [18] to interpret framework prediction. A global interpreter SHAP is used over LIME [22], to interpret the effect of the single fea- ture on the target variable. SHAP framework utilizes vari- ous explainability methods for better interpretation of model prediction. The proposed methodology is divided into three major stages i) Ensemble learning, ii) Prediction of learning, iii) Sign Explainer, and interpret the results. Figure 2 shows the sequential flow of the proposed model.

ENSEMBLE LEARNING

Every custom Deep Learning model is based on training- based learning and must necessarily stage to learn deep features. Especially, when the task was related to computer vision, proper model training is necessary. The proposed methodology used ensemble learning with an attention model.

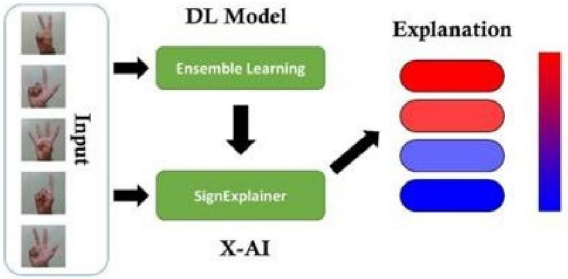

FIGURE 2. Sequential process architecture of proposed methodology.

Figure 3 represents an ensemble attention-based model for sign language recognition. The proposed method- ology uses a bagging-based ensemble model to learn the associated feature of sign images. Attention-based Ensem- ble learning mainly divides into two categories, multi- head ensemble and attention-based ensemble [23]. Figure 3 demonstrate the different way of attention-based ensemble learning. Algorithm 1 represents the architectural structure of the proposed ensemble learning approach with the bagging concept.

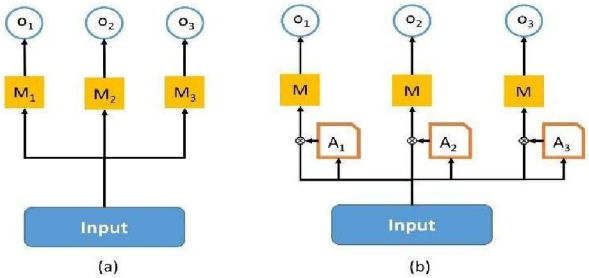

FIGURE 3. A different way to use attention for ensemble learning,

represents a multi-head ensemble with different feature embedding parameters, represents the same feature embedding with different feature learning

The proposed methodology used ensemble learning. Which is mainly divided into two parts. The first one is ResNet50 with a 23.521M parameter as part of the con- volution learning module. ResNet50 is used to reduce the vanishing gradient problem. Generally, in a deep convolution network loss function is shrunk to zero after several itera- tions. With the help of the ResNet network, gradients can be directed to skip connections from previous layers to the next filter layer. The linear learning of residual network can be considered as equation 1. [24], where *G (x, and {Wi})* stand for mapping of residual learning, while W_S and x stand for projection square matrix of x dimension.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

$\eta = G(x, \{W_i\}) + W_S + x$ (1)

Another component of ensemble learning is the attention module, which can be designed with two associated modules as feature extraction module F(x) and attention module A(x). The feature extraction module was designed with a pro-layer perceptron model, and generalized as equation 2 [23]. And the attention weights were calculated as equations 3 and 4, where he and hd stand for encoder and decoder weights.

$F(x) = h_i(h_{ii}-1(\ldots(h2(h1(x)))$ (2)

$\gamma = \tanh(W * he + W * hd)$ (3)

$A(x) = Softmax(\gamma)$ (4)

Algorithm 1 Pseudo-Code for Proposed Ensemble Learning Architecture (Bagging based)

Input: Training Image set *I*
Output: Interpretation Index
1. $K \leftarrow Conv\_Layer(ResNet(i))$
2. $l \leftarrow Class\_Labels\ \{0,1,2\ldots,A,B,\ldots Z\}$
3. $G \leftarrow Ensemble\_feature(l)$
4. $C \leftarrow Num\_Classes(l)$
5. **for** $k \in \{1, \ldots, K\}$ **do**
6. **for** $C \in \{0, \ldots, C\}$ **do**
7. $D_c = Conv(I_{c0} * I_{c1} * \ldots\ldots * I_{cn})$
8. $f_c = D_0 \cup D_1 \cup \ldots\ldots\ldots \cup D_n$
9. **end for**
10.
11. **end for**
12. $G(x) = softmax((G1(x) + G2(x) + \ldots\ldots + Gk(x)) / K)$
13. #Feature Explainer:
14. **procedure** Sign_Ex(g(x), *l*)
15. $i \leftarrow max\_val(int)$
16. $Create\overline{\Pi}\ for\ collections$
17. **for** i ∈ g(x) **do**
18. **for each** $\pi \in \{\pi_0 \ldots\ldots\ldots \pi_1\}$ **do**
19. $Calculate\pi_i;$
20. $\pi_0 \leftarrow \Delta(\pi_i)$
21. **end for**
22. $Y \leftarrow$ evaluate(π_0, *l*)
23 **end for**
24. **return**(index ← max_val(Y))
25. **end procedure**

The global feature embedding model $G(x)$ (equa- tion 5), for the embedding module. Authors have proposed three-dimension blob channel to recognize input images in an RGB channel. The attention feature and convolution feature are associated with the final feature v(e1c)tor generation and it was forwarded to a fully connected DCNN network for classification. Figure 4 represents the conceptual architecture representation of the proposed ensemble learning with the attention model.

$G(x) = F(x) \otimes A(x)$ (5)

CLASSIFICATION AND PREDICTION

The output from the fully connected layer is further pro- cesses for classification and prediction. The authors have implemented multi-layer perceptron (MLP) [25] to classify sign language. The proposed methodology uses DFFN (Deep Forward Neural Network) to recognize gesture signs from input images. ReLU activation was implemented in the final layer of the deep network for sign recognition, and it can be calculated as equation (6), where ($W1$, $W2$) are different weights and ($b1$, $b2$) as bias.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

$DFNN = ReLU\ (W1x + b1)\ W2 + b2$ (6)

The authors have utilized NumPy and Scikit-learn [26]
for evaluation and visualization. The class-wise performance score has been calculated, and accuracy, precision, recall, and F1-Score were calculated to analyze ensemble model performance. Performance standards have been calculated as per equations 7 to 10. [27], [28].

$Accuracy = (TP + TN)/(TP + FP + TN + FN)$ (7)

$Precision = TP\ (TP + FP)$ (8)

$Recall = TP\ (TP + FN)$ (9)

$F\ 1 - Score = 2*(Precision*Recall)/(Precision+Recall)$ (10)

SIGN EXPLAINER

Interpretation and explainable techniques involved with black-box deep learning models fall under two categories, model specific or agnostic. This section focuses on the design of SignExplainer an agnostic interpretability technique, that can be applied to any black-box deep-learning model to interpret gesture-based signs. SHAP [29] is among the most utilized interpretability methods for deep learning-based methods. SHAP can construct interpretations for multi-class classifier responses. SignExplainer uses Sign- specific Xcon- cept to generate a fault line explanation. Let's assume that δ_{pred} and δ_{alt} can be Xconcept for E_{alt} and E_{alt} respectively where E stands for the actual class. Based on Xconcept, line prediction can be calculated as equation 11 [30].

$$W(E_{pred}, E_{alt}) \leftarrow \min_{\delta_{pred}, \delta_{alt}} \alpha\ \delta_{pred}, \delta_{alt}\ + \beta\ |\delta_{pred}|\ + \lambda\ ||\delta_{alt}|| \quad (11)$$

The proposed Methodology designs DeepExplainer as an additive feature attribution method with accuracy and miss-ingness. DeepExplainer combines the SHAP value computed for a smaller component of the ensemble network and calculates it as equation 12, [31]. Where, $o = f(x) - f(r)$ and $x_i = x_i - x_r$, $r$is the reference input, while $f(x)$ is the model output.

$$O = \sum_{i=1}^{n} Cx_i * \Delta o \quad (12)$$

IV. EXPERIMENTS AND RESULTS

DATASET

The authors have evaluated SignExplainer with ensemble learning on Indian Sign Language Dataset [32]. The dataset used for simulation consists of 36 Indian Sign classes having digits (0-9) and an alphabet (A-Z). The dataset consists of approximately 1200 images per class, with 3 channel images. Along with Indian Sign Language (ISL) dataset, the authors have also experimented with other static datasets like Amer- ican Sign Language (ASL) [33], and Bangla Sign Language (BSL) [34]. Property of datasets described in Table 2.

TABLE 2. Statistical representation of different sign language datasets used in the simulation.

Dataset	Avg. Resolution	Classis	Avg. Image per class
Indian Sign Language (ISL) [32]	250 x 250	36	1200
American Sign Language (ASL) [33]	400 x 400	35	840
Bangla Sign Language (BSL) [34]	171 x 166	33	654

DATA AUGMENTATION

The proposed simulation uses data augmentation to make the model more generalized for feature learning. Data augmentation is also used to balance training image samples and improve robustness for learning variability over the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

different images, making the model more generalized toward real- time scenarios. Direct image inference may yield biased findings due to particular transformations and noise associated with equipment and surroundings. Image augmentation must be used to achieve more reliable and robust prediction to improve accuracy and prevent overfitting. The authors have implemented i) Geometric transformations as random horizontal flip,
random rotation with +0.2 to -0.2, and zooming by 1.5% to 2.5%. ii) Color space transformations as random RGB change and Brightness by 0.5%. Figure 5 represents the sample of the augmented training dataset.

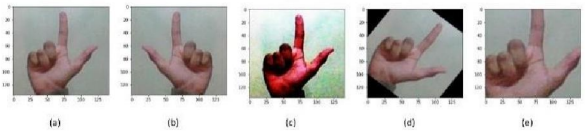

FIGURE 4. Input Sign image augmentation, (a) original image, (b) horizontal flip, (c) color transformation, (d) random rotation, (e) zooming.

SIMULATION DETAILS

The authors have implemented training of an ensemble learn- ing module on the ISL dataset [32]. TensoFlow- Keras has been used for the design of the proposed methodology.
The proposed ensemble methodology has achieved 98.20 % accuracy with extracted features from attention and the ResNet50 model. Model training was divided with 0.2 train- test split ratios (80:20) for all experiments, with an image size of (72, 72, 3) and a batch size of 16. The model was simulated with 0.3 as a dropout ratio and a 0.001 learning rate with the Adam optimizer. Table 3 demonstrate superior performance over other standard Convolution networks, additionally, the best performance was observed by the proposed Attention- based ensemble model. The proposed methodology has achieved significant accuracy over 50 learning epochs, as shown in Figure 6.

INTERPRETATION WITH SIGNEXPLAINER

The proposed methodology simulates SignExplainer to gen- erate a model prediction and explain the correctness of the prediction. The simulation uses OpenCV for masking the input images and passes them to the ‘‘blur (128,128)’’ method, which is responsible to mask the predicted image output with inpainttelea value. The authors have created SignExplainer with adaptive feature abstraction, which com- pares with and without x- features. X-

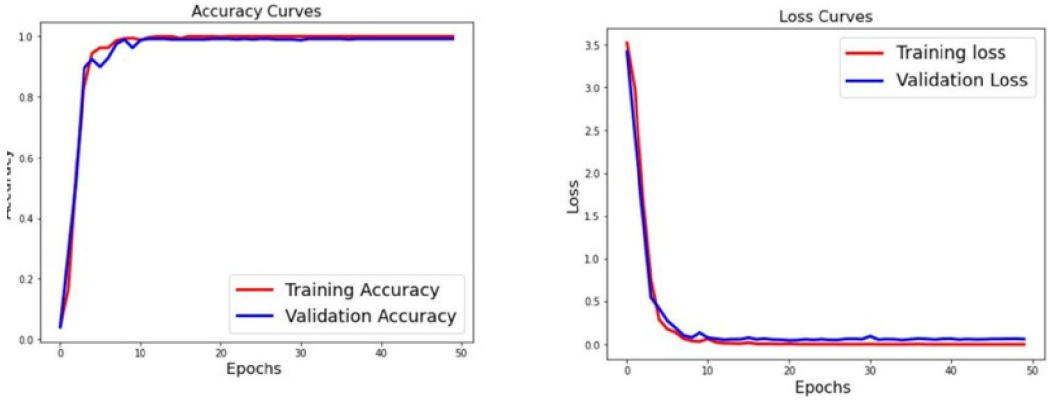

FIGURE 5. Accuracy and loss curve for Indian Sign Language recognition using Attention-based Ensemble learning.
Features are the associative contribution of ensemble learning features. Prediction function of SignExplainer, which is working as a masked feature. The authors have passed sign images with the Explainer object to generate SHAP values, and Figure 6 represents the plot.
The interpretation plot has been taken with 4 flips over 1,000 evaluations as (max_eval=1000) for the Explainer object (shown in figure 7). The gradient bar prediction represents the prediction’s relevance interpretation, red stands for the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

maximum, and blue stands for the minimum. Table 4 represents the performance of other basic XAI models to interpret the ensemble model prediction output over the Indian Sign Language dataset.

TABLE 3. Performance analysis with state-of-the-art models for Image classification

Model	Accuracy	Precision	Recall	F1-Score
CNN [35]	92.60%	0.92	0.92	0.91
VGG16 [36]	97.55%	0.98	0.97	0.97
EfficientNet V2 [37]	96.42%	0.96	0.96	0.95
Ensemble (ResNet50 + Attention)	98.20%	0.98	0.98	0.97

We have demonstrated the remarkable result of explanation over sign language, especially for Indian signs. To ensure the robustness of the proposed SignExplainer with an ensem- ble learning model, the author has evaluated the proposed methodology over other static and standard sign language datasets like American Sign Language ASL) and Bangla Sign Language (BSL), statistical comparison describe in Table 5. The prediction score of SignExplainer for the test sign image is demonstrated in Figure 8. SignExplainer helps to understand and recognizes why the model recognizes the data instance as it has. The first image is from the testing dataset as a significant gesture of ‘‘4.’’ The top of all predictions shows the matching value. Red dots represent high relevance while Blue dots represent low relevance. Based on the high relevance of feature attribution it’s easy to interpret how the model was learned to predict sign language. The presence of a red pixel over the corresponding area of the hand gesture increases the prediction probabilities.

TABLE 4. Statistical performance comparison of different models for Interpretation over ISL dataset (where TRP is True Positive Rate, FNR is False Negative Rate, PPV is Positive Predictive value, and FDR is False Discovery Rate).

Explainer Model	TRP	FNR	PPV	FDR
DeepLine [38]	80.8	19.2	70.8	29.2
Lime [21]	82.4	17.6	72.9	27.1
DeepLIFT [39]	79.1	20.9	74.8	25.2
SignExplainer	87.3	12.7	78.5	21.5

V. RESULTS ANALYSIS AND DISCUSSION

A Computer vision-based model to learn and interpret the prediction was proposed by this study. The authors have proposed a sequential (two-phase) methodology from learn- ing from the ensemble model to interpretation of the pre- dicted result, with the SignExplainer model. The authors have also implemented the proposed architecture for Indian Sign Language (ISL). Experiment also extends to other static sign languages like American Sign Language (ASL), Bangla Sign Language. This study proposed and demon- strated attention-based ensemble learning with ResNet50 and Self-attention model. The proposed architecture was able to achieve 98.20 percent remarkable accuracy for ISL, and also compare with other computer vision state-of-the- art models. The second phase of the study demonstrated the interpretation of the learning model. The authors have used the SignEx- plainer model to extract masked values from the black-box model.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

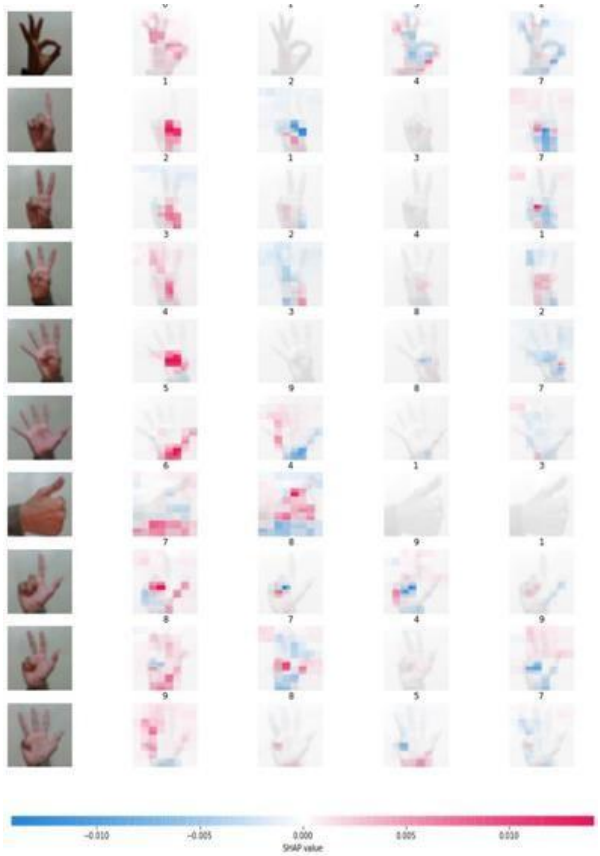

FIGURE 6. Support feature for SignExplainer over Indian Sign Language Recognition, (a few samples have been taken to maintain article readability).

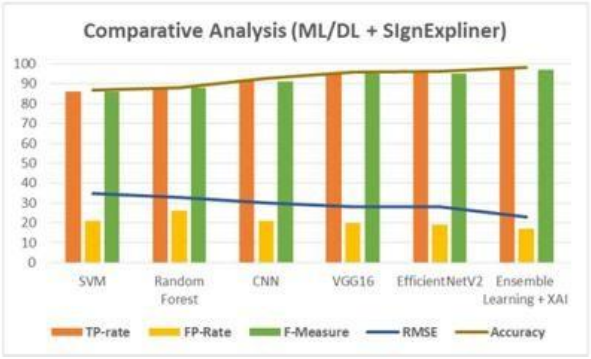

FIGURE 7. Support feature for SignExplainer over Indian Sign Language Recognition, (a few samples have been taken to maintain article readability).

The proposed SignExplainer uses fault line calculation to interpret the correctness of the predicted sign image. The result section also demonstrates the achieved result by Sign- Explainer, and also compare it with other conventional XAI model. The author has also evaluated TP-rate and FP-rate for the proposed model, and it's found remarkable with other black box deep learning models as 0.98 and 0.17 respectively. Figure 9 represents a comparative analysis of the proposed architecture

(ensemble learning + SignExplainer) with other deep learning models like SVM [40], Random Forest [41], CNN [35], VGG16 [36], and EfficientNetV2 [37]. The evaluation matrix was calculated with a True-False positive rate, F-measures, and RMSE (Root Mean Square Error) value. The statistical analysis represents the proposed associative

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

architecture is more accurate than other standard machine learning and deep learning models (shown in Figure 9). The authors have also analyzed other deep learning object detection models like R-CNN [42], Faster R-CNN [43], and Single Shot Detector (SSD) [44] with VGG16 [45] as the backbone over the proposed Attention-based Ensemble model. A comparative analysis of deep learning detection models was illustrated in Figure 10. Figure 11 illustrates the confusion matrix of the proposed ensemble learning methodology for the static Indian Sign Language dataset

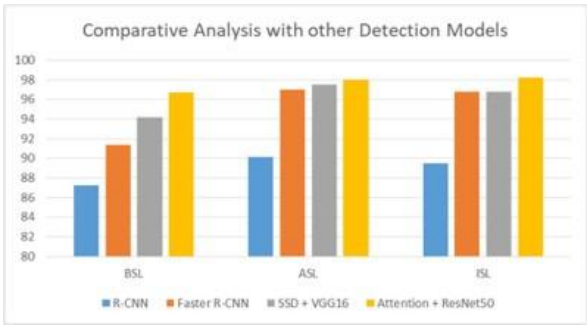

FIGURE 8. Comparative accuracy analysis of proposed methodology

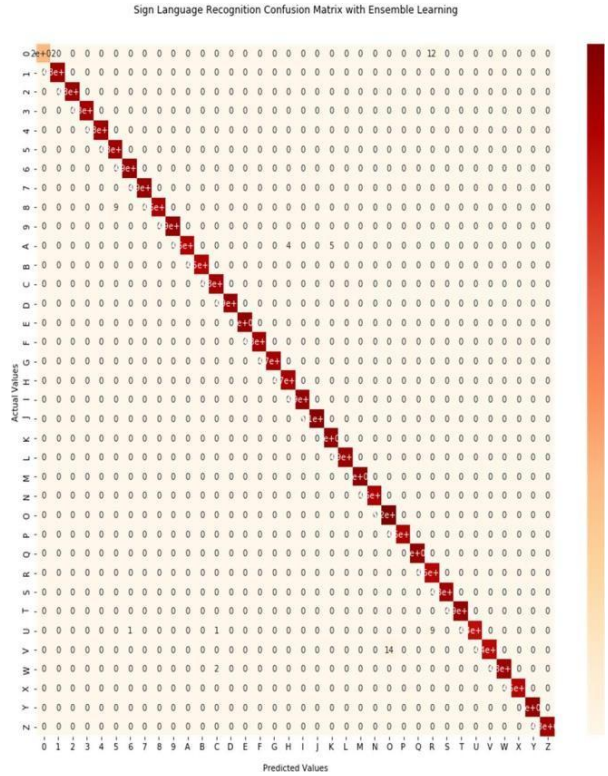

FIGURE 9. Confusion Matrix for Static Indian Sign Language using Ensemble Learning with ResNer50.

VI. CONCLUSION

The era of Explainable AI growing exponentially, to over- come trust and transparency issues of deep learning models. Especially tasks relevant to Computer vision or NLP must require interpreting predicted results over critical sectors. The review has explored different XAl methodologies like LRP, LIME, SHAP, and SmoothGrad over relevant computer vision applications. This study has proposed Sign Language Recognition to make explainable artificial intelligence. Ensemble learning-based architecture was proposed to recognize sign gestures from sign images. Ensemble weights were passed to the proposed SignExplainer to generate statistical values like TP-rate and FP-rate, to

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

evaluate the correctness of the proposed SignExplainer. This study also evaluated ensemble learning with another deep learning model for image classification. The proposed study also evaluates the performance of SignExplainer over other benchmark static sign language datasets like ASL and BSL, and it also achieves remarkable performance. The proposed study also simulates additional machine learning and deep learning models like Decision tree, Random Forest, VGG16, and EfficientNetV2, and evaluates the performance of SignExplainer. Ensemble learning and other deep learning models were also per- formed well over SignExplainer to interpret predicted signs with proper statistical values. The proposed work can be extended to other static Sign Languages as well as isolated Sign Languages. The proposed methodology can be enhanced for real-time or portable Sign Language Recognition with acceptable interpretations.

ACKNOWLEDGMENT

This research was funded by Princess Nourah bint Abdulrah- man University and Researchers Supporting Project number (PNURSP2023R346), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors would also like to acknowledge the support of Prince Sultan University for paying the Article Processing Charges (APC) of this publication

REFERENCES

[1] dataset for the ground truth evaluation of neural network explanations,'' Inf. Fusion, vol. 81, pp. 14–40, May 2022, doi: 10.1016/j.inffus.2021. 11.008.

[2] P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, and P. M. Atkinson, [19] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, ''Explainable artificial intelligence: An analytical review,'' WIREs Data Mining Knowl. Discovery, vol. 11, no. 5, p. e1424, 2021.

[3] Y. Yuan and Y. Lo, ''Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks,'' IEEE J. Biomed. Health Informat., vol. 23, no. 2, pp. 519–526, Mar. 2019, doi: 10.1109/jbhi.2017.2787487.

[4] A. Gramegna and P. Giudici, ''SHAP and LIME: An evaluation of dis- criminative power in credit risk,'' Frontiers Artif. Intell., vol. 4, Sep. 2021, Art. no. 752558.

[5] F. Afza, M. A. Khan, M. Sharif, S. Kadry, G. Manogaran, T. Saba, Ashraf, and R. Damaševičius, ''A framework of human action recogni- tion using length control features fusion and weighted entropy-variances based feature selection,'' Image Vis. Comput., vol. 106, Feb. 2021, Art. no. 104090.

[6] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, ''Explainable AI: A review of machine learning interpretability methods,'' Entropy, vol. 23, no. 1, p. 18, Dec. 2020, doi: 10.3390/e23010018.-

[7] M. Baldeon Calisto and S. K. Lai-Yuen, ''AdaEn-net: An ensemble of adaptive 2D–3D fully convolutional networks for medical image segmentation,'' Neural Netw., vol. 126, pp. 76–94, Jun. 2020, doi: 10.1016/j.neunet.2020.03.007.

[8] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, ''DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,'' IEEE Trans. Pat- tern Anal. Mach. Intell., vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.

[9] J. Ganesan, A. T. Azar, S. Alsenan, N. A. Kamal, B. Qureshi, and A. E. Hassanien, ''Deep learning reader for visually impaired,'' Electron- ics, vol. 11, no. 20, p. 3335, Oct. 2022.

[10] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-González, and J. M. Corchado, ''Deepsign: Sign language detection and recognition using deep learning,'' Electronics, vol. 11, no. 11, p. 1780, Jun. 2022, doi: 10.3390/electronics11111780.

[11] B. Kim, M. Wattenberg, J. Gilmer, C. Cai, J. Wexler, F. Viegas, and R. Sayres, ''Interpretability beyond feature attribution: Quantita- tive testing with concept activation vectors (TCAV),'' in Proc. Int. Conf. Mach. Learn., Mar. 2023, pp. 2668–2677. [Online]. Available: http://proceedings.mlr.press/v80/kim18d.html

[12] A. Dhurandhar, P.-Y. Chen, R. Luss, C.-C. Tu, P. Ting, K. Shanmugam, and P. Das, ''Explanations based on the missing: Towards contrastive explanations with pertinent negatives,'' 2018, arXiv:1802.07623.

[13] A. Akula, S. Wang, and S.-C. Zhu, ''CoCoX: Generating concep- tual and counterfactual explanations via fault-lines,'' in Proc. AAAI Conf. Artif. Intell., Apr. 2020, vol. 34, no. 3, pp. 2594–2601, doi: 10.1609/aaai.v34i03.5643.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

[14] V. Contreras, N. Marini, L. Fanda, G. Manzo, Y. Mualla, J.-P. Calbimonte, M. Schumacher, and D. Calvaresi, ‘‘A DEXiRE for extracting propositional rules from neural networks via binarization,’’ Electronics, vol. 11, no. 24, p. 4171, Dec. 2022, doi: 10.3390/electronics11244171.
[15] J. Patel, C. Amipara, T. A. Ahanger, K. Ladhva, R. K. Gupta, H. O. Alsaab, Y. S. Althobaiti, and R. Ratna, ‘‘A machine learning-based water potability prediction model by using synthetic minority oversampling technique and explainable AI,’’ Comput. Intell. Neurosci., vol. 2022, pp. 1–15, Sep. 2022, doi: 10.1155/2022/9283293.
[16] T. Vermeire, D. Brughmans, S. Goethals, R. M. B. de Oliveira, and D. Martens, ‘‘Explainable image classification with evidence counterfac- tual,’’ Pattern Anal. Appl., vol. 25, no. 2, pp. 315–335, Jan. 2022, doi: 10.1007/s10044-021-01055-y.
[17] Y. Goyal, Z. Wu, J. Ernst, D. Batra, D. Parikh, and S. Lee, ‘‘Coun- terfactual visual explanations,’’ in Proc. 36th Int. Conf. Mach. Learn., May 2019, pp. 2376–2384, Accessed: Mar. 2023. [Online]. Available: https://proceedings.mlr.press/v97/goyal19a.html
[18] L. Arras, A. Osman, and W. Samek, ‘‘CLEVR-XAI: A benchmark ‘‘Learning deep features for discriminative localization,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2921–2929, Accessed: Mar. 6, 2023. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2016/html/Zhou_Learning_Deep_Features_CVPR_2016_paper.html
[20] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, ‘‘Grad-CAM: Visual explanations from deep networks via gradient-based localization,’’ Int. J. Comput. Vis., vol. 128, no. 2, pp. 336–359, Feb. 2020, doi: 10.1007/s11263-019- 01228-7.
[21] M. T. Ribeiro, S. Singh, and C. Guestrin, ‘‘Why should i trust you?: Explaining the predictions of any classifier’’ in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 1135– 1144.
[22] X. Shen, K. Lu, S. Mehta, J. Zhang, W. Liu, J. Fan, and Z. Zha, ‘‘MKEL: Multiple kernel ensemble learning via unified ensemble loss for image classification,’’ ACM Trans. Intell. Syst. Technol., vol. 12, no. 4, pp. 1–21, Aug. 2021.
[23] W. Kim, B. Goyal, K. Chawla, J. Lee, and K. Kwon, ‘‘Attention-based ensemble for deep metric learning,’’ in Proc. Eur. Conf. Comput. Vis. (ECCV, 2018, pp. 736–751
[24] B. Chen and W. Deng, ‘‘Deep embedding learning with adaptive large margin N-pair loss for image retrieval and clustering,’’ Pattern Recognit., vol. 93, pp. 353–364, Sep. 2019, doi: 10.1016/j.patcog.2019.05.011.
[25] D. R. Kothadiya, C. M. Bhatt, T. Saba, A. Rehman, and S. A. Bahaj, ‘‘SIGNFORMER: DeepVision transformer for sign language recognition,’’ IEEE Access, vol. 11, pp. 4730–4739, 2023, doi: 10.1109/access.2022.3231130.
[26] J. Mueller and L. Massaron, Python for Data Science. Hoboken, NJ, USA:Wiley, 2019.
[27] J. Huang, W. Zhou, H. Li, and W. Li, ‘‘Sign language recognition using real-sense,’’ in Proc. IEEE China Summit Int. Conf. Signal Inf. Process. (ChinaSIP, Jul. 2015, pp. 166–170.
[28] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, ‘‘Sign lan- guage recognition using convolutional neural networks,’’ in Proc. Eur. Conf. Comput. Vis., 2015, pp. 572–578.
[29] S. Knapič, A. Malhi, R. Saluja, and K. Främling, ‘‘Explainable artificial intelligence for human decision support system in the medical domain,’’ Mach. Learn. Knowl. Extraction, vol. 3, no. 3, pp. 740–770, Sep. 2021, doi: 10.3390/make3030037.
[30] J. van der Waa, E. Nieuwburg, A. Cremers, and M. Neerincx, ‘‘Evaluating XAI: A comparison of rule-based and example-based explanations,’’ Artif. Intell., vol. 291, Feb. 2021, Art. no. 103404.
[31] F. Gabbay, S. Bar-Lev, O. Montano, and N. Hadad, ‘‘A LIME- based explainable machine learning model for predicting the severity level of COVID-19 diagnosed patients,’’ Appl. Sci., vol. 11, no. 21, p. 10417, Nov. 2021.
[32] D. R. Kothadiya. (Oct. 2022). Deepkothadiya/STATIC_ISL: Static Indian Sign Language Dataset Having Sign of Digit and Alphabet. [Online]. Available: https://github.com/DeepKothadiya/Static_ISL
[33] Thakur. (May 2019). American Sign Language Dataset. [Online]. Available: https://www.kaggle.com/datasets/ayuraj/american- sign- language-dataset
[34] S. M. Rayeed. (Aug. 2021). Bangla Sign Language Dataset. [Online]. Available: https://www.kaggle.com/datasets/rayeed045/bangla-sign- language-dataset

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, April 2024

[35] T. Saba, M. A. Khan, A. Rehman, and S. L. Marie-Sainte, ‘‘Region extraction and classification of skin cancer: A heterogeneous framework of deep CNN features fusion and reduction,’’ J. Med. Syst., vol. 43, no. 9, Jul. 2019, doi: 10.1007/s10916-019-1413-3.
[36] K. Simonyan and A. Zisserman, ‘‘Very deep convolutional networks for large-scale image recognition,’’ 2014, arXiv:1409.1556.
[37] B. Li, B. Liu, S. Li, and H. Liu, ‘‘An improved EfficientNet for Rice germ integrity classification and recognition,’’ Agriculture, vol. 12, no. 6, p. 863, Jun. 2022, doi: 10.3390/agriculture12060863.
[38] Y. Heffetz, R. Vainshtein, G. Katz, and L. Rokach, ‘‘DeepLine: AutoML tool for pipelines generation using deep reinforcement learning and hierarchical actions filtering,’’ in Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2020, pp. 2103–2113.
[39] H. Chen, S. Lundberg, and S.-I. Lee, ‘‘Explaining models by propagat- ing Shapley values of local components,’’ in Explainable AI in Health- care and Medicine. Cham, Switzerland: Springer, 2020, pp. 261–270. [Online]. Available: https://link.springer.com/book/10.1007/978-3-030- 53352-6?page=2#toc, doi: 10.1007/978-3-030-53352-6.
[40] A. Razaque, M. Ben Haj Frej, M. Almi’ani, M. Alotaibi, and B. Alotaibi, ‘‘Improved support vector machine enabled radial basis function and linear variants for remote sensing image classification,’’ Sensors, vol. 21, no. 13, p. 4431, Jun. 2021, doi: 10.3390/s21134431.
[41] Z. Noshad, N. Javaid, T. Saba, Z. Wadud, M. Saleem, M. Alzahrani, and O. Sheta, ‘‘Fault detection in wireless sensor networks through the random forest classifier,’’ Sensors, vol. 19, no. 7, p. 1568, Apr. 2019.
[42] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, ‘‘Oriented R-CNN for object detection,’’ 2021, arXiv:2108.05699.
[43] Y. Liu, ‘‘An improved faster R-CNN for object detection,’’ in Proc. 11th Int. Symp. Comput. Intell. Design (ISCID), vol. 2, Dec. 2018, pp. 119–123.
[44] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, ‘‘SSD: Single shot multibox detector,’’ Proc. Comput. Vis. (ECCV), 2016, pp. 21–37.
[45] A. T. Azar, Z. I. Khan, S. U. Amin, and K. M. Fouad, ‘‘Hybrid global optimization algorithm for feature selection,’’ Comput., Mater. Continua, vol. 74, no. 1, pp. 2021–2037, 2023.