IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 15 August 2024, accepted 30 September 2024, date of publication 4 October 2024, date of current version 31 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3474011

RESEARCH ARTICLE

KUNet-An Optimized AI Based Bengali Sign Language Translator for Hearing Impaired and Non Verbal People

ABDULLAH AL JAID JIM[1,2], IBRAHIM RAFI[2], JUN JIAT TIANG[3], UZZAL BISWAS[2], AND ABDULLAH-AL NAHID[2]

[1]Department of Electrical and Electronics Engineering, Trust University, Barishal 8200, Bangladesh
[2]Electronics and Communication Engineering Discipline, Khulna University, Khulna 9208, Bangladesh
[3]Centre for Wireless Technology (CWT), Faculty of Engineering, Multimedia University, Cyberjaya 63100, Malaysia

Corresponding authors: Jun Jiat Tiang (jjtiang@mmu.edu.my) and Abdullah-Al Nahid (nahid.ece.ku@gmail.com)

This work was supported by the Research Management Centre, Multimedia University, Malaysia.

ABSTRACT Sign language is the most prevalent form of communication among people with speech and hearing disabilities. The most widely used types of sign language involve the creation of static or dynamic gestures using hand(s). Among many sign languages, Bengali Sign Language (BdSL) is one of the most complicated sign languages to learn and comprehend because of its enormous alphabet, vocabulary, and variation in expression techniques. Existing solutions include learning BdSL or hiring an interpreter. Besides, BdSL interpreter support is hard to come by and expensive (if not voluntary). Disabled people might find it more comfortable to converse with generals implementing machine translation of sign language. Deep learning that mimics the human brain, a subset of the machine learning domain, seems to be a viable solution. For the hearing impaired and non verbal community, computer vision, in particular, may hold the key to finding a solution. Therefore, we have created a novel model, KUNet ("Khulna University Network" a CNN based model), a classification framework optimized by the genetic algorithm (GA), has been proposed to classify BdSL. This model and the dataset contribute to creating a BdSL machine translator. GA-optimized KUNet acquired an accuracy of 99.11% on KU-BdSL. After training the model on KU-BdSL, we demonstrated a comparison of the model with state-of-the-art studies and interpreted the black-box nature of the model using explainable AI (XAI). Additionally, we have found that our model outperformed several well-known models trained on the KU-BdSL dataset. This study will benefit the hearing impaired and non verbal community by allowing them to communicate effortlessly and minimizing their hardship.

INDEX TERMS Bengali sign language (BdSL), classification, computer vision, deep learning, machine learning, sign language recognition.

I. INTRODUCTION

Every day we do many easy tasks as part of our daily routine, like brushing our teeth or talking to someone. Yet, for many people, these simple tasks take immense effort due to disability. A person with an abnormal physical or mental condition that limits the natural ability of a human being is considered disabled, and his condition is termed disability. Some people are born with disabilities, while others may develop these conditions later in life. Irrespective of the development period, disabled people face many difficulties in specific circumstances. There are broadly four categories of disabilities [1], namely - hearing disability, vision disability, and cognitive disability. A person may have one disability or may have several. World Health Organization (WHO) estimates that 1 billion people (15% of the world's population) are affected by a physical, sensory (such as hearing impaired or vision impaired), intellectual, or mental health disability that significantly interferes with their everyday existence [2]. Among these large populations, speech and hearing-impaired people face numerous challenges when communicating with the general people. We practically all encounter people like

The associate editor coordinating the review of this manuscript and approving it for publication was Jeon Gwanggil.

them at least once in our lives. We want to help them as much as they want assistance from us. The only barrier is the communication medium, which is primarily sign language. Being so essential for this enormous community, it lacks universality. It varies from one country to another. Again, it can have diversity within a geographical boundary, for instance, Bengali sign language (BdSL). The two broad groups of BdSL are - static sign language [3], [4], [5], and dynamic sign language [6], [7], [44]. These two groups can further differ depending on the involvement of the number of hands. However, due to the complexity of double-handed gestures, single-handed ones were used for BdSL alphabet representation.

The existing solutions for understanding speech and hearing impaired people is to learn sign language or to seek help from someone who has learned. Both take courage and effort to learn sign language, which are the primary obstacles to learning sign language. Moreover, patience and practice are required to learn and achieve expertise. In comparison to human learning, computers can learn quite quickly. Hence, we can solve the dilemma of learning sign language by adopting deep learning. A large population in Bangladesh is suffering from the curse of hearing impairment, resulting in around 2.4 million BdSL users in Bangladesh, according to [8]. Being essential for this large community, the BdSL datasets are underdeveloped. The state-of-the-art BdSL datasets are not as advanced as other sign languages, such as American sign language (ASL) datasets [9] or British sign language (BSL) datasets [10].

We have worked on BdSL to contribute towards the betterment of the hearing impaired and non verbal people, which involves creating a novel deep learning model named KUNet, a classification framework optimized by the GA that identifies the hand gestures. The model is comparatively shallow in nature, as it has a simple architecture with a few layers. A shallow model has a few trainable parameters [11], a few layers [12], simple construction, and easy interpretability. Also, a shallow model requires low time complexity and space complexity [13]. Hence, the KUNet model can reduce the computational time [14] (represented in section IV). The optimization part of the model has been accomplished by a GA. To achieve better accuracy, the GA chose the model's hyperparameters. The model attained better accuracy, yet we were using a model that lacks interpretability. The use of a black-box model could limit the contribution towards the D&D society. Even newer regulations (for instance, European General Data Protection Regulation (GDPR) [15]) restrict the adoption of black-box models. Thus, in order to maintain the accessibility and usability of this research, we have used explainable AI (XAI) to handle sensitive data and interpret the model. XAI demonstrates the pixels responsible for the prediction. We can decide whether to trust the model's prediction from the XAI results. The principal contributions of the research are illustrated below.

- Recognition of the BdSL consonant using KUNet that might alleviate the difficulty of learning BdSL by providing ML-based translation.
- Exploration of the utilization of GA optimization in BdSL recognition, which is yet to be discovered.
- Use of XAI for interpretation of the outputs to enhance machine learning (ML) based translation of BdSL.

The sole purpose of this study is to ease the communication with the D&D people and to do so, we have organized the rest of the article as follow. In section II, we have provided contemporary research article knowledge that is further subdivided into four research groups. Section III includes all the necessary procedures to achieve our goal. GA optimization result, comparison with related works and state-of-the-art model, and XAI result are presented in section IV. The conclusion of this research and a brief future research prospect is present in section V.

II. LITERATURE REVIEW

We have collected the relevant research articles from the Web of Science (WOS) repository using the keywords "Bangla sign language" and "Bengali sign language". We have chosen the research articles published between the years 2014 to 2023 (till 13 August 2023). Filtering the papers, in accordance with the years, we have 55 papers in total. After checking all of them, we eliminated 12, which did not match our purpose. We have categorized the remaining 43 research articles into two groups. The categories are based on the number of hands used in presenting a BdSL hand sign. The first group is double-handed BdSL, and the other group is single-handed BdSL. Table 1 portrays the groups of the articles found according to the number of hands used.

TABLE 1. Illustration of the categorized groups and their corresponding research articles.

Group	Reference
Double-handed BdSL	[7], [8], [16] - [38]
Single-handed BdSL	[39] - [56]

A. DOUBLE-HANDED BdSL

In double-handed BdSL, a person requires both hands to represent a hand sign. Each hand gesture corresponds to a number, alphabet, or word. We reviewed several double-handed BdSL in this study. Most researchers involve double-handed BdSL while working with both numerals and alphabet.

One of the primitive works on double-handed BdSL is [23], in which Yasir et al. recognized 15 Bengali consonants. They made a dataset by taking samples from 22 different people. They have employed Linear Discriminant Analysis (LDA) along with an Artificial Neural Network (ANN) model to serve the classification purpose.

The researches like [7] and [27] have made the advancement of machine learning based BdSL recognition. They worked on both vowels and consonants of BdSL. Uddin and Chowdhury [7] recognized hand signs using principal component analysis (PCA) and support vector machine (SVM) models, where a dataset incorporating 4800 images of 480 × 360 pixels yield an accuracy of 97.70% on the segmentation task. On the other hand, [27] utilized the Ishara-Lipi database [29], on which their CNN model acquired a precision of 99.86%.

The researches like [7] and [27] have made the advancement of machine learning based BdSL recognition, in which both vowels and consonants of BdSL are recognized. Uddin and Chowdhury [7] recognized hand signs using principal component analysis (PCA) and support vector machine (SVM) models, where a dataset incorporating 4800 images of 480×360 pixels yield an accuracy of 97.70% on the segmentation task. On the other hand, [27] utilized the Ishara-Lipi database [29], on which their CNN model acquired a precision of 99.86%.

Word level BdSL requires the use of both hands. Both [25] and [32] worked on Bengali sign language on a word level. In [25], they trained a system utilizing a dataset comprising 27 classes, each having 20 images. Among these 27 classes, 16 classes represent 16 different words, and the rest are Bengali numerals. Ten volunteers participated in the curation of this dataset. They also claimed they required no wearable gloves or markings for their work. Their research achieved an accuracy of 96.46%.

B. SINGLE-HANDED BdSL

In the single-handed BdSL, the user utilizes one of the hands to express their thoughts. Single-handed BdSL is way easier than double-hand BdSL. As a result, recently, single-handed BdSL is gaining popularity. Hence, we have also selected the single-handed BdSL approach for this research work. In this category, we discovered 18 relevant research articles that can be further divided into subcategories depending on hand sign recognition. We split the articles into four groups and gave them the names Groups A, B, C, and D. The BdSL numbers, BdSL consonants, BdSL alphabet (vowels and consonants), and BdSL words are progressively recognized by Groups A, B, C, and D. An overview of these groups is presented in Table 2.

TABLE 2. Description of different groups of single-handed BdSL research articles.

Group	Reference	Recognition	Recognition Example
A	[39] - [42]	BdSL digits	০, ৯
B	[43] - [45]	BdSL consonants	ক, খ
C	[46] - [55]	BdSL alphabet	আ, ক
D	[56]	BdSL words	সাহায্য, চাই

1) GROUP A RESEARCH ARTICLES

Among the single-handed BdSL works, [39], [40], [41], [42] belong to Group A research articles that recognize Bengali numerals from hand signs. All of these researches detect 0-9 with satisfactory recognition rates. A complete dataset for Bengali digits (0 to 9) is presented using MediaPipe, a cross-platform depth-map estimation framework [39]. Here, the necessary depth information is obtained by the hand skeleton joints from the RGB images. Among various classifiers, SVM has reached the highest accuracy on the proposed dataset, which is 98.65%. Hasan et al. [40] used the histogram of oriented gradients (HOG) model to classify the BdSL digits. For this noble work, they utilized the Ishara-lipi dataset [29]. Though Ishara-lipi is a double-handed dataset, its BdSL digits are collected using the single-hand approach. Reference [40] obtained an accuracy of 94.74% in its classification task. A rule-based system for Bengali voice and text to BdSL numeral interpretation is proposed by [41]. Creating a dataset of 1674 images from various participants in different conditions, [42] employed Convolution Neural Network (CNN) to interpret BdSL digits.

2) GROUP B RESEARCH ARTICLES

Group B research articles focus on the recognition of BdSL consonants. Among this group, one of the earliest studies of single-handed BdSL representing consonants, Ayshee et al. [43] created a fuzzy rule-based technique for recognizing static single expressions for only two single-hand BdSL alphabet. Calculating finger angles, essentially fuzzy criteria for matching with established rules for two Bengali letters, determine the hand configurations. Only a single perspective can be utilized to input a static image of a single hand gesture in this scheme. One disadvantage of this technology is that it is impossible to see all of the fingers from a single perspective. In [44], Ahmed et al. proposed a self-made dataset of 518 images representing 37 classes. They classified the Bengali consonants with 98.99% accuracy with an ANN. Uddin et al. [45] also categorized Bengali consonants from their self-made dataset using the SVM model with 86% accuracy.

3) GROUP C RESEARCH ARTICLES

Group C research articles include [46], [47], [48], [49], [50], [51], [52], [53], [54], and [55] that detect BdSL vowels and consonants. Two approaches, conventional transfer learning and contemporary zero-shot learning (ZSL), for automatic BdSL letter detection, are presented in [46]. Their model achieved harmonic mean accuracy, seen accuracy, and zero-shot accuracy of 68.21%, 91.57%, and 54.34%, respectively, with six unseen classes. The pre-trained DenseNet201 architecture was the most effective feature extractor for the transfer learning-based approach, according to the researchers. After completing quantitative experimentation on 18 CNN architectures and 21 classifiers, Linear Discriminant Analysis (LDA) becomes the best classifier, with an overall accuracy of 93.68% on the massive dataset. Ref. [47] adopts dynamic skin calibration and geometric hashing to interpret the BdSL letters. The overall recognition rate is relatively low for [47], which is 51.35%. However, the authors prepared a new database of 1147 images for training a hash table map with geometric coordinates of the feature points. Shanta et al. [48] combined scale-invariant feature transform (SIFT) feature extraction with CNN for the first time for hand sign classification. There were 38 classes representing 51 alphabet, each having 200 images for this classification task. Reference [49] has an overall detection rate of 96.33% on the training dataset and 84.68% on the validation dataset of [47]. Employing the obtained features from a pre-trained model and fine-tuned top layers of Deep CNN yields [49] a better recognition rate. A two-way translation (normal text to sign language and vice versa) is presented in [50]. They produced a dataset of 2050 photos for this work, which were classified using a custom CNN. They achieved accuracy as high as 92.85%.

In conjunction with the Bangladesh National Federation of the Deaf, [51] currently comprises the largest BdSL dataset, with 12,581 distinct hand signs corresponding to 38 different BdSL alphabet. They gathered the samples from 320 participants, including 42 hearing impaired individuals. Rafi et al. [51] have utilized VGG19 architecture for the BdSL alphabet recognition, and 224 $\times$ 224 pixel images were selected as input to this model. They obtained an overall accuracy of 89.6%. Ref. [52] chose a deep CNN model and 37 letters (8 vowels and 29 consonants) for contribution to BdSL. They collected 3219 images from 6 individuals in the creation of the dataset. A data glove to translate the hand signs requiring movement of the hand had been introduced by [53]. These dynamic hand gestures cannot be recognized by employing an image dataset. Sayeed et al. [54] utilized a modified InceptionV3 model and transfer learning technique to acquire an accuracy of 94.41% on the [51] dataset. Urmee et al. [55] has developed the BdSLInfinite dataset, which has been employed to train a CNN model using Xception architecture. They obtained an accuracy of 98.93% with an average response time of 48.53ms.

4) GROUP D RESEARCH ARTICLES

Working on the alphabet level is tiresome while expressing a word using any sign language. Hence, word-based representation of sign language is essential. Ref. [21], a word-based BdSL study employed a self-made dataset of BdSL words. In this work, the authors applied SVM to identify the Bengali word of the corresponding hand gesture. They achieved a precision of 86.53%. As SVM alone will not yield the desired accuracy, they incorporated HOG with SVM. Also, they added audio output after detecting the signs.

III. METHODOLOGY

In this research, we introduced a novel KUNet classification framework. From the three variants of the KU-BdSL dataset [57], we have used the Uni-scale Sign Language Dataset (USLD) for this work. We have optimized the KUNet model by introducing a GA. GA took the basic structure of our model (a sequence of layers and their interconnections) as inputs and produced an optimal model based on accuracy by varying the hyperparameters. Later, this GA-optimized model was used for classification tasks, and we compared our findings with other state-of-the-art researches on BdSL.

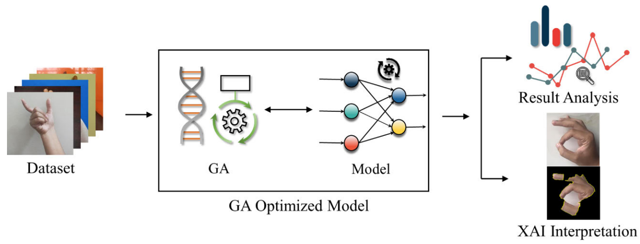


FIGURE 1. **Overall methodology of the research.**

We also compared our model with commonly used models (trained on KU-BdSL), namely VGG16 [58], VGG19 [58], AlexNet [59], and ResNet50 [60]. Finally, we used the Local Interpretable Model-agnostic Explanations (LIME) as the XAI method to interpret and increase trustworthiness of the model. An overview of our methodology is portrayed in FIGURE 1.

A. DATASET

The dataset comprised 30 classes, each having 50 images totaling 1500 images. USLD images have dimensions of 512 $\times$ 512 pixels, which would slows down the GA process. Therefore, in order to reduce the processing cost and time, we downsampled the images to 64 $\times$ 64 pixels. The 30 classes resemble 38 consonants ('banjonborno') of Bengali alphabet. The number of consonants in BdSL surpassed the number of classes, as some hand gestures demonstrate more than one letter. We augmented the images by flipping (vertically and horizontally), rotating, and varying the contrast by 50%, 20%, and 20% of the total samples, respectively. The dataset was split into train data (70%) for training the KUNet and test data (30%).

B. GENETIC ALGORITHM (GA)

In 1988, Goldberg and Holland introduced the GA [61] for machine learning mimicking biological evolution. Since then, GA has been applied for constrained and unconstrained optimization problems [62], [63]. In this study, we have utilized the GA to optimize the layer parameters of the KUNet (layers are described in section III-C). Table 3 illustrates which parameters of the layers were optimized using the GA and what were the ranges of optimization.

TABLE 3. **Parameters that are optimized in different layers.**

Sl. No	Layer	Parameter	Range of Parameter
1	Conv2D	No. of Filters, Kernel-size	32 to 2048 (multiple of 32), 1 to 5
2	Maxpooling	Pool-size	2 to 3
3	Dropout	Rate	5% to 50%
4	Activation	Function	ReLU [64] or ELU [65] or Softmax
5	Dense	Units	32 to 4096 (multiple of 32)

We started with 200 populations, which were random variations of the model. The populations are depicted as

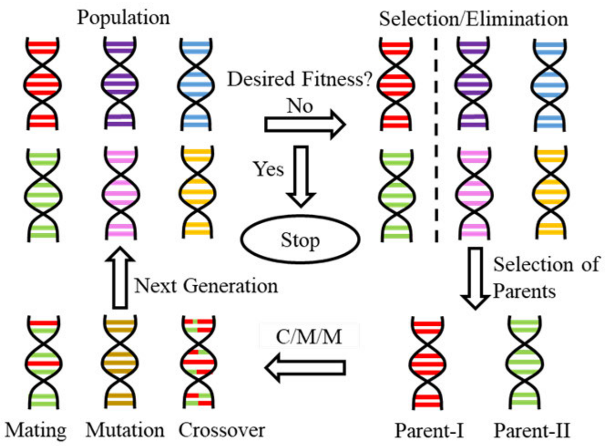


FIGURE 2. Our genetic algorithm.

DNAs, and the base pairs represent layers in FIGURE 2. The models were trained with 25 epochs to reduce the processing cost, and the accuracies were calculated for their respective layer parameters. Then the models were sorted in declining order by their accuracy. The top 10% of the population with the best accuracy were immediately elected as the next-generation population, while the least performing 50% models were discarded. The remaining units (40%) and the already elected top 10% participated in generating the remaining 90% of its initial population for the next generation. Two entities (Parent-I and Parent-II) were chosen and subjected to the GA operations at random from these populations, which entails three processes: crossover, mating, and mutation (C/M/M).

In mating, we took all the parameters present in a layer from either of the parents randomly. We have selected different parameters for a single layer from different parents while employing the crossover. In contrast, the mutation process generated a completely distinct population with separate parameters from the parents to avoid local maxima and local minima. The individual layer parameter of a population has been optimized by applying one of the C/M/M processes. These processes continue producing a new population for the next-generation until the total count reaches 90% of its initial population. There was a 90% likelihood (each parent has 45% chance) that parameters in a layer come exclusively from either parent-I or parent-II during the mating process. There was a 5% probability each for crossover and mutation. All the recently generated populations with the best performing 10% of the previous generation produced the new generation. These populations further progressed through fitness calculation, and the rest of the process continued until they reached the desired accuracy. The GA process was terminated whenever the accuracy was adequate.

C. NOVEL OPTIMAL KUNet CLASSIFICATION FRAMEWORK

We proposed a novel KUNet classification framework for Bengali sign language recognition in this study, and

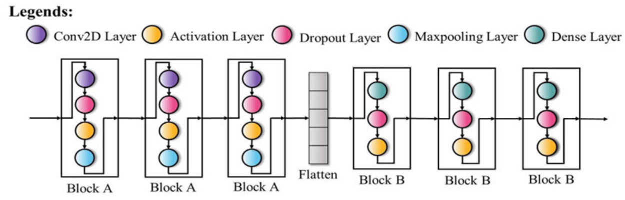


FIGURE 3. Novel KUNet classification framework.

the model includes conv2D, maxpooling, dense, dropout, activation layers, flatten, and batch normalization. FIGURE 3 demonstrates the novel KUNet structure depicted using the combination of two blocks, Block A and Block B, for a simple representation of KUNet. Block A consists of conv2D, dropout, activation, and maxpooling layers, and Block B involves dense, dropout, and activation layers.

The conv2D layer output $\mathcal{G}[m, n]$ was generated by equation 1, where $\mathcal{I}_{img}$ and $\mathcal{K}$ denotes input images and kernels, and i, j are iterable parameters.

$$\begin{aligned}\mathcal{G}[m, n] &= (\mathcal{I}_{img} * \mathcal{K})[m, n] \\ &= \sum_i \sum_j \mathcal{K}[i, j]\mathcal{I}_{img}[m-i, n-j]\end{aligned} \tag{1}$$

The dropout layer dumps interconnections between neurons at random, which helps to evade the co-adaptation of feature detectors [66]. Maxpooling is introduced to reduce computational cost by downsampling an input feature map. Rectified Linear Units (ReLU) and Exponential Linear Units (ELU) have been utilized for activation, which respectively follows equation 2 and equation 3. In equation 2 and 3, $f(x)$ represents ReLU or ELU output, where x = node for dense layers and pixel for conv2D layers. GA determines the suitable activation function for each layer in the model.

$$f(x) = max(0, x) \tag{2}$$

$$f(x) = \begin{cases} x & \text{if x > 0} \\ a\exp(x) - 1 & \text{otherwise} \end{cases} \tag{3}$$

A flatten layer is employed to convert two-dimensional inputs into single-column output after the deployment of three Block A's. This output is then used as a source of information for dense layers. There are 3200 weighted neurons (units) in our GA optimized KUNet, each neuron is fully connected to its previous layer's neurons in the dense layer. The dense layer performs a matrix-vector multiplication on the input vectors, as indicated in equation 4. The symbols $\mathcal{X}_j$, $\mathcal{W}_{ji}$, and $\mathcal{b}$ denote neurons, weight, and bias, respectively, which are employed in a matrix-vector multiplication to produce the dense layer output ($\mathcal{Z}_i$).

$$\mathcal{Z}_i = \sum_j \mathcal{W}_{ji}\mathcal{X}_j + \mathcal{b} \tag{4}$$

Batch normalization (BN) is introduced in both Block A and B (BN follows equation 5-8) to lower the image pixel values, reducing computational complexity. Here, μ_m, σ_m, and $\hat{Z}$ sequentially represents mean, standard deviation, and the final output. Again, $Z^{(i)}$ represents activation vector, which is the output of an activation layer, and $Z_{norm}^{(i)}$ is its normalized version. Finally, two trainable parameters γ and β are adopted to determine the BN output, applying a linear transformation.

$$\mu_m = \frac{1}{n}\sum_i Z^{(i)} \quad (5)$$

$$\sigma_m = \frac{1}{n}\sum_i (Z^{(i)} - \mu_m) \quad (6)$$

$$Z_{norm}^{(i)} = \frac{Z^{(i)} - \mu_m}{\sqrt{\sigma^2 - \epsilon}} \quad (7)$$

$$\hat{Z} = \gamma * Z_{norm}^{(i)} + \beta \quad (8)$$

KUNet accepts 64×64 pixel images as input and produces all the 30 class probabilities of the image. The class having the highest probability is then selected as the label of the image.

D. EXPLAINABLE AI (XAI)

Explainable AI (XAI), also known as Interpretable AI, is a form of artificial intelligence (AI) that enables people to comprehend ML models. Among the different XAI methods, we used Local Interpretable Model Agnostic Explanation (LIME) [67], a popular model-agnostic XAI method that works with images. The term 'local fidelity' refers to the requirement for the explanation to accurately reflect the classifier's behavior 'around' the instance being predicted. Lime is model-agnostic in that it can explain any model without having to 'peek' into it.

By feeding datasets into a machine learning model, LIME analyses the outcomes and interprets the data. When LIME is given a data sample as input, it adjusts various data features. LIME then keeps a record of these input variations in order to create a new local dataset. Utilizing this newly generated dataset to train a surrogate model, weighted by the sampled instances' proximity to the instance of interest. Finally, LIME estimates the relevance of input features using the local dataset and surrogate model. For images, LIME changes individual pixels randomly, which could have little effect on the predictions. As a result, picture variants are formed by segmenting the image into superpixels and turning them on or off. Superpixels are linked pixels with similar hues that may be disabled by changing each pixel's color to a user-defined color (in this study, black). Additionally, the user may define a chance for turning off a superpixel in each permutation.

IV. RESULT

The total parameters of the ML models employed in this research have been portrayed in Table 4. Our KUNet has the second-lowest total number of parameters after ResNet50. AlexNet has 5,183,328 more parameters than KUNet, making it the next-lowest total number of parameters. VGG16 and VGG19 models have more than double the number of total parameters compared to KUNet.

TABLE 4. Number of parameters of the ML models.

Model Name	Total Parameters	Trainable Parameters	Non-trainable Parameters
KUNet	57,246,934	57,234,330	12,604
AlexNet [59]	62,430,262	62,419,318	10,944
ResNet50 [60]	23,649,182	23,596,062	53,120
VGG16 [58]	134,383,454	134,383,454	0
VGG19 [58]	139,693,150	139,693,150	0

Table 5 exhibits minimum, maximum, mean, and standard deviation times to predict 15 samples where the models are trained using 100 iterations. The mean time (μ_t) and the standard deviation time (σ_t) have been determined using eq. 9 and eq. 10. In the equation, we have applied N = number of samples = 15 and $Parameter$ = Time to predict a sample in milliseconds. KUNet requires the second least amount of time to predict 15 samples (AlexNet takes the least time). On average, to classify 15 samples, the KUNet model requires 132.9 ms longer than the AlexNet model, and among the remaining models, ResNet50 takes the minimum time (average) that is nearly 4.5 times longer, despite having the lowest number of total parameters. Compared to KUNet, VGG16 and VGG19 require roughly six and seven times longer, respectively. KUNet has the second-lowest (2.40ms) standard deviation after AlexNet (2.07ms). ResNet50 and VGG16 have almost twice the standard deviation as KUNet, whereas VGG19 has eleven times the standard deviation.

$$\mu = \frac{\sum_N Parameter}{N} \quad (9)$$

$$\sigma = \sqrt{\frac{\sum_N (Parameter - \mu)^2}{N}} \quad (10)$$

TABLE 5. Number of parameters of the ML models.

Model Name	Minimum Time (ms)	Maximum Time (ms)	Mean Time (ms)	Standard Deviation (ms)
KUNet	279.42	299.24	283.28	2.40
AlexNet [59]	146.21	159.60	150.38	2.07
ResNet50 [60]	1271.33	1294.41	1279.78	4.57
VGG16 [58]	1715.71	1815.96	1736.11	26.77
VGG19 [58]	2093.60	2121.20	2106.64	4.29

We have designed the basic structure of the novel KUNet model and let the GA optimize it. GA performed remarkably in terms of network optimization. As portrayed in FIGURE 4, the model improved its accuracy (or fitness) with generations. FIGURE 4 exhibits the accuracy of the best-performing network among the 200 populations or networks for each generation. Initially, when we generated 200 populations (generation-0) by randomly varying the parameters of our KUNet model, the accuracy of the best-performing network was 93.78%. Then, from the first to the fifteenth generation, the accuracy increased significantly and reached 99.78%. The accuracy remained the same for the next 19 generations.

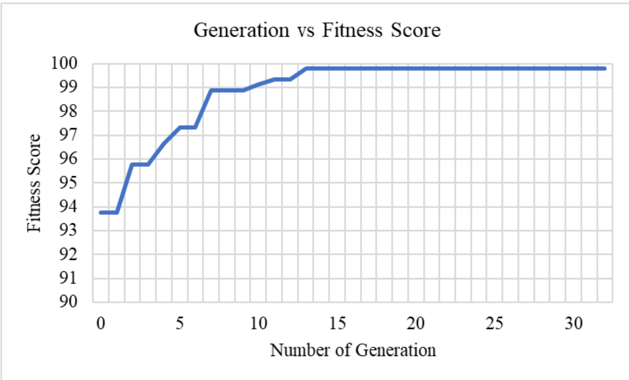


FIGURE 4. Accuracy of the GA optimized model at individual generation.

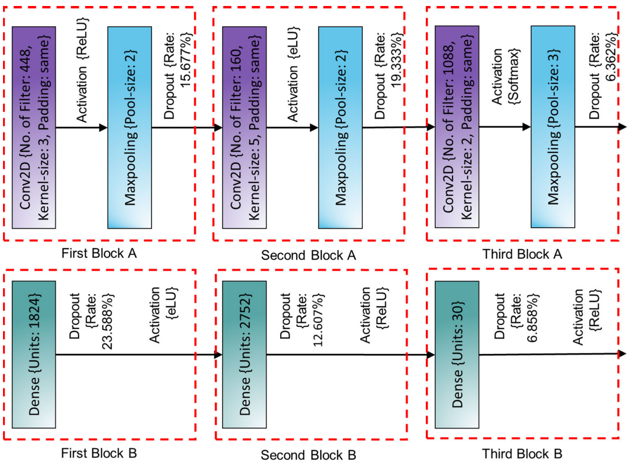


FIGURE 5. Summary of the KUNet.

We took the hyperparameters of the best-performing network of generation-22 and created the novel KUNet classification framework. The summary of the optimal KUNet classification framework is portrayed in FIGURE 5. The optimal values for each layer are exhibited for blocks A and B, as mentioned in the methodology section.

The accuracy and loss during training of the KUNet are depicted in FIGURE 6. The validation accuracy is greater than the training accuracy because of dropout layers used in the network. For loss calculation, we have utilized sparse categorical cross-entropy loss. Regarding loss and accuracy, test cases perform better than train cases. This phenomenon is explainable due to the dropout.

FIGURE 7 illustrates the confusion matrix of the model. The classes represented in the axis of the confusion matrix correspond to BdSL alphabets. The confusion matrix verifies that other than some exceptions, our model predicts all classes quite accurately. Most of the classes have a 100 percent recognition rate. Only class 4 has one false detection among 450 test cases (30% of the total dataset). The model has predicted it as a sample of the 2454 class (predicted class) of the KUBdSL dataset instead of the 2454 class (actual class).

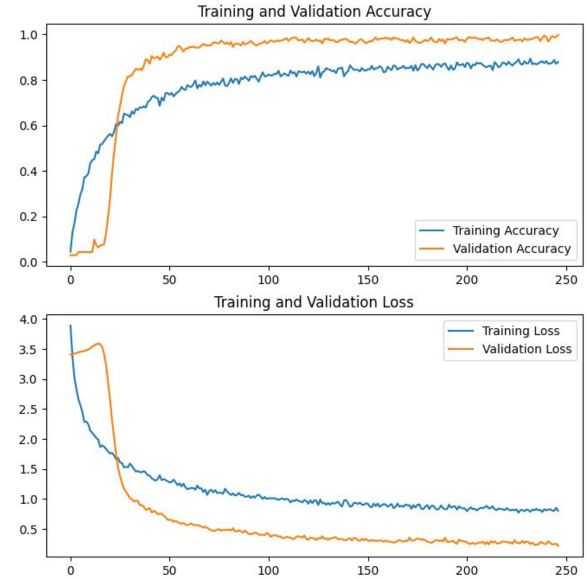


FIGURE 6. Summary of the KUNet.

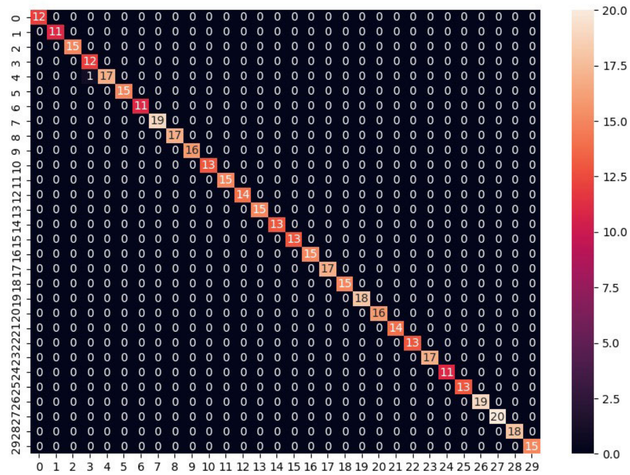


FIGURE 7. Summary of the KUNet.

We trained some of the renowned deep learning models (for example- AlexNet [59], ResNet50 [60], VGG16 [58], and VGG19 [58]) on the KU-BdSL dataset and evaluated them. Table 6 illustrates that the novel KUNet classification framework outperforms all these models classifying KU-BdSL images. Among the renowned models, the AlexNet is the best-performing model, with 0.01 percent less accuracy, 0.09 percent greater precision, 0.05 percent greater F-measure, and similar recall. Based on the information in Table 6, we may conclude that our model significantly achieves better performance with a simple model structure and better tuning. All this information is graphically presented in FIGURE 8 for comparison.

BdSL has many approaches for expressing the same letters, hence the comparison with related works becomes

TABLE 6. Comparison with other State-of-the-art models.

Model Name	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
KUNet	99.78	99.74	99.81	99.77
AlexNet [59]	99.77	99.83	99.81	99.82
ResNet50 [60]	62.22	64.35	63.42	61.24
VGG16 [58]	99.77	99.83	99.80	99.81
VGG19 [58]	99.56	99.34	99.17	99.20

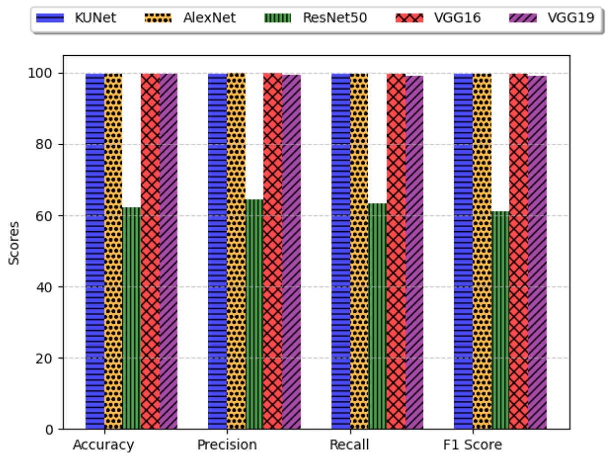


FIGURE 8. Performance comparison with other models.

TABLE 7. Comparison with related works.

Work	Model Architecture	GA Optimization	Dataset	Accuracy
[51]	VGG19	No	[51]	89.60%
[68]	Efficient NNE	No	[68]	93.00%
[69]	RNN	No	[69]	98.19%
			[51]	93.80%
KUNet	Novel Optimal KUNet Framework	No	[70]	99.78%

complicated. However, as we selected the single-handed BdSL approach, we have compared our work with other state-of-the-art single-handed technique research, for instance, [51], [68], and [69]. We have trained and tested our model on the KU-BdSL dataset and exhibited the findings of other researchers in Table 7. From Table 7, we observe that our novel KUNet classification framework has obtained an accuracy of 99.78% on the KU-BdSL dataset and outperforms all the other studies' accuracy. Our model has obtained 10.17%, 6.77%, and 1.58% more accuracy than that of [51], [68], and [69], respectively, when tested on the self-created dataset. Reference [69] reached an accuracy of 93.8% on [51]'s dataset, which is 5.97% less than ours. Also, Table 7 shows that only our model is GA optimized, which yields us the best results.

To have a better understanding of the performances of the utilized models on the KUBdSL dataset, we have calculated the mean and standard deviation of the models' accuracy,

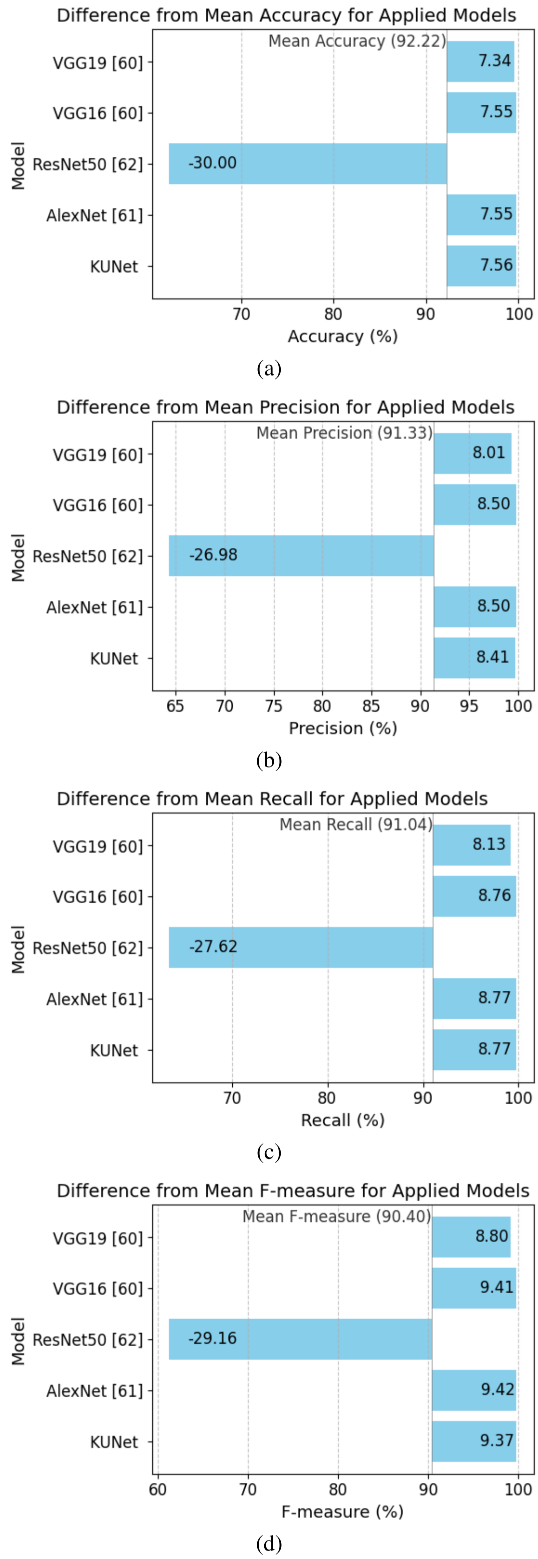


FIGURE 9. Difference between mean value and (a) Accuracy, (b) Precision, (c) Recall, and (d) F-measure of the applied models.

precision, recall, and F-measure (inspired by in study [71]) and exhibited in Table 8. The mean and standard deviation of the evaluation parameters are determined using eq. 9 and

TABLE 8. Mean and standard deviation of the evaluation parameters of the applied ML models on the KUBdSL dataset.

SL. No.	Parameters	Mean	Standard Deviation
1	Accuracy	92.22	15.00
2	Precision	91.33	14.19
3	Recall	91.04	14.56
4	F-measure	90.40	15.45

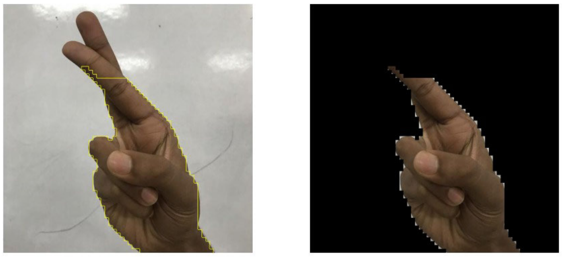


FIGURE 10. Class 2480/2524/2525 image sample and responsible features.

FIGURE 11. Explanation of a misclassified sample from class 2466 by LIME.

eq. 10, respectively; where, N is the total number of models ($N = 5$; as five models are applied in this study to classify the KUBdSL dataset) and parameter denotes accuracy, precision, recall, or F-measure.

FIGURE 9 illustrates the plots of the difference between different parameters of the utilized models and their mean values. FIGURE 9a and FIGURE 9c exhibit that the KUNet has the best accuracy and recall (AlexNet has matched the recall result) compared to their mean values of the models. VGG16 and AlexNet have the same difference in the mean accuracy and precision (as demonstrated in FIGURE 9a and FIGURE 9b) FIGURE 9 displays all the models perform above average in all scenarios except the ResNet50 model.

A. XAI RESULT

From XAI, we have illustrated the relevant portion of the images responsible for any single image falling into a particular class by running the model through the LIME method. For instance, FIGURE 10 is a sample image from class 2480/2524/2525 (which is predicted correctly), and the XAI result represents the region (super-pixels) that is responsible for its predicted class. It demonstrates that the model has recognized the hand sign, evaluating the correct pixels.

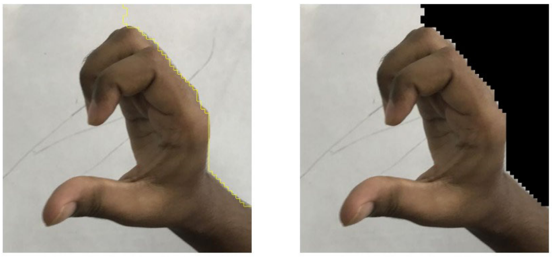


FIGURE 12. Explanation of a misclassified sample from class 2454 by LIME.

In contrast to the correct prediction, the model failed to recognize FIGURE 11 and 12. In both cases, the model could not perform accurately learning the features. For FIGURE 11, the model found no feature at all. However, for FIGURE 12, the model partially learned the features of the image, which caused misclassification.

V. CONCLUSION

The importance of sign language cannot be overstated. Deep learning can provide the required assistance. This technology is a cost-effective and efficient solution to the hardship of these people. Our novel KUNet classification framework is one of the many solutions from the deep learning sector, which achieved an accuracy of 99.78%. The model surpasses most of the popular deep learning architectures (e.g., AlexNet, ResNet50, VGG16, VGG19) in terms of accuracy, recall, precision, and F1 score. The GA based optimization of the model is what accounts for this level of operational excellence. Also, GA optimization is a one-time process that results in a reliable and accurate model. Another important aspect of our research is that we used XAI to explain our black-box model. As the issue of language is sensitive towards D&D people, it was necessary to clarify that the model does not produce wrong decisions after getting new data. We have successfully addressed the problem, and the XAI result supports our statement. After the accomplishment of our research, we hope to minimize the struggle and hardship of hearing and speech impaired people. Our study has substantial social implications as well as contributions to the research community. It introduces genetic algorithms with deep learning models to the BdSL and explains the black-box nature of deep learning models using explainable AI. This is the first work on Bengali sign language recognition that uses a GA and explainable AI, and the possibilities are endless.

The computational cost of utilizing the GA increases as the number of training samples increases, which is a drawback of the study. For better performance, it requires massive computational cost. Hence, we will analyze whether increasing the computational cost improves accuracy in the future. Also, we will be creating a BdSL dataset of 49 Bengali letters (both vowels and consonants) with a large number of sample images. Our model is insufficient for real-time application, as it can only predict a single hand gesture from an input image. However, practical scenarios might include multiple hand gestures in sample images. Therefore, we will be focusing on a real-time application-based model, which will identify all the gestures present in an image. We will additionally optimize the layers as well as the layer parameters.

REFERENCES

[1] W. Horner-Johnson, K. Dobbertin, J. C. Lee, and E. M. Andresen, ''Disparities in chronic conditions and health status by type of disability,'' *Disability Health J.*, vol. 6, no. 4, pp. 280–286, Oct. 2013.

[2] *World Report on Disability*, World Health Org., Geneva, Switzerland, 2011.

[3] K. Deb, M. I. Khan, H. P. Mony, and S. Chowdhury, ''Two-handed sign language recognition for Bangla character using normalized cross correlation,'' *Global J. Comput. Sci. Technol.*, vol. 12, no. 3, pp. 1–6, 2012.

[4] D. S. H. Pavel, T. Mustafiz, A. I. Sarkar, and M. Rokonuzzaman, ''Geometrical model based hand gesture recognition for interpreting Bengali sign language using computer vision,'' in *Proc. ICCIT*, 2003, pp. 642–647.

[5] B. Sarkar, K. Datta, C. D. Datta, D. Sarkar, S. J. Dutta, I. D. Roy, A. Paul, J. U. Molla, and A. Paul, ''A translator for Bangla text to sign language,'' in *Proc. Annu. IEEE India Conf.*, Oct. 2009, pp. 1–4.

[6] F. M. Rahim, T. E. Mursalin, and N. Sultana, ''Intelligent sign language verification system using image processing, clustering and neural network concepts,'' *Int. J. Eng. Comput. Sci. Math.*, vol. 1, no. 1, pp. 43–56, 2010.

[7] M. A. Uddin and S. A. Chowdhury, ''Hand sign language recognition for Bangla alphabet using support vector machine,'' in *Proc. Int. Conf. Innov. Sci., Eng. Technol. (ICISET)*, Oct. 2016, pp. 1–4.

[8] M. T. Hoque, M. Rifat-Ut-Tauwab, M. F. Kabir, F. Sarker, M. N. Huda, and K. Abdullah-Al-Mamun, ''Automated Bangla sign language translation system: Prospects, limitations and applications,'' in *Proc. 5th Int. Conf. Informat., Electron. Vis. (ICIEV)*, May 2016, pp. 856–862.

[9] D. Li, C. R. Opazo, X. Yu, and H. Li, ''Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison,'' in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 1448–1458.

[10] S. Albanie, G. Varol, L. Momeni, H. Bull, T. Afouras, H. Chowdhury, N. Fox, B. Woll, R. Cooper, A. McParland, and A. Zisserman, ''BBC-oxford British sign language dataset,'' 2021, *arXiv:2111.03635*.

[11] C. Li, G. Chen, G. Liang, and Z. Y. Dong, ''A novel high-performance deep learning framework for load recognition: Deep-shallow model based on fast backpropagation,'' *IEEE Trans. Power Syst.*, vol. 37, no. 3, pp. 1718–1729, May 2022.

[12] K. Pasupa and W. Sunhem, ''A comparison between shallow and deep architecture classifiers on small dataset,'' in *Proc. 8th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, Oct. 2016, pp. 1–6.

[13] F. Lei, X. Liu, Q. Dai, and B. W.-K. Ling, ''Shallow convolutional neural network for image classification,'' *Social Netw. Appl. Sci.*, vol. 2, no. 1, p. 97, Jan. 2020.

[14] G. R. B. Ferreira and H. V. H. Ayala, ''Improved image-based welding status recognition with dimensionality reduction and shallow learning,'' *Experim. Mech.*, vol. 62, no. 6, pp. 985–998, Jul. 2022.

[15] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, ''What do we need to build explainable AI systems for the medical domain?'' 2017, *arXiv:1712.09923*.

[16] D. Talukder, F. Jahara, S. Barua, and M. M. Haque, ''OkkhorNama: BdSL image dataset for real time object detection algorithms,'' in *Proc. IEEE Region 10 Symp. (TENSYMP)*, Aug. 2021, pp. 1–6.

[17] S. N. Hasan, M. J. Hasan, and K. S. Alam, ''Shongket: A comprehensive and multipurpose dataset for Bangla sign language detection,'' in *Proc. Int. Conf. Electron., Commun. Inf. Technol. (ICECIT)*, Sep. 2021, pp. 1–4.

[18] N. Begum, R. Rahman, N. Jahan, S. S. Khan, T. Helaly, A. Haque, and N. Khatun, ''Borno-net: A real-time Bengali sign-character detection and sentence generation system using quantized YOLOv4-tiny and LSTMs,'' *Appl. Sci.*, vol. 13, no. 9, p. 5219, Apr. 2023.

[19] S. K. Youme, T. A. Chowdhury, H. Ahamed, M. S. Abid, L. Chowdhury, and N. Mohammed, ''Generalization of Bangla sign language recognition using angular loss functions,'' *IEEE Access*, vol. 9, pp. 165351–165365, 2021.

[20] S. Das, M. S. Imtiaz, N. H. Neom, N. Siddique, and H. Wang, ''A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier,'' *Exp. Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 118914.

[21] K. K. Podder, M. E. H. Chowdhury, A. M. Tahir, Z. B. Mahbub, A. Khandakar, M. S. Hossain, and M. A. Kadir, ''Bangla sign language (BdSL) alphabets and numerals classification using a deep learning model,'' *Sensors*, vol. 22, no. 2, p. 574, Jan. 2022.

[22] T. Tazalli, Z. A. Aunshu, S. S. Liya, M. Hossain, Z. Mehjabeen, M. S. Ahmed, and M. I. Hossain, ''Computer vision-based Bengali sign language to text generation,'' in *Proc. IEEE 5th Int. Conf. Image Process. Appl. Syst. (IPAS)*, Dec. 2022, pp. 1–6.

[23] R. Yasir and R. A. Khan, ''Two-handed hand gesture recognition for Bangla sign language using LDA and ANN,'' in *Proc. 8th Int. Conf. Softw., Knowl., Inf. Manage. Appl. (SKIMA)*, Dec. 2014, pp. 1–5.

[24] M. A. Rahaman, M. Jasim, M. H. Ali, and M. Hasanuzzaman, ''Computer vision based Bengali sign words recognition using contour analysis,'' in *Proc. 18th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2015, pp. 335–340.

[25] S. M. K. Hasan and M. Ahmad, ''A new approach of sign language recognition system for bilingual users,'' in *Proc. Int. Conf. Electr. Electron. Eng. (ICEEE)*, Nov. 2015, pp. 33–36.

[26] F. Yasir, P. W. C. Prasad, A. Alsadoon, and A. Elchouemi, ''SIFT based approach on Bangla sign language recognition,'' in *Proc. IEEE 8th Int. Workshop Comput. Intell. Appl. (IWCIA)*, Nov. 2015, pp. 35–39.

[27] M. J. Hossein and M. Sabbir Ejaz, ''Recognition of Bengali sign language using novel deep convolutional neural network,'' in *Proc. 2nd Int. Conf. Sustain. Technol. Ind. 4.0 (STI)*, Dec. 2020, pp. 1–5.

[28] M. M. Hasan, M. Khaliluzzaman, S. A. Himel, and R. T. Chowdhury, ''Hand sign language recognition for Bangla alphabet based on Freeman chain code and ANN,'' in *Proc. 4th Int. Conf. Adv. Electr. Eng. (ICAEE)*, Sep. 2017, pp. 749–753.

[29] M. S. Islam, S. S. S. Mousumi, N. A. Jessan, A. S. A. Rabby, and S. A. Hossain, ''Ishara-lipi: The first complete MultipurposeOpen access dataset of isolated characters for Bangla sign language,'' in *Proc. Int. Conf. Bangla Speech Lang. Process. (ICBSLP)*, Sep. 2018, pp. 1–4.

[30] S. Islam, S. S. S. Mousumi, A. S. A. Rabby, S. A. Hossain, and S. Abujar, ''A potent model to recognize Bangla sign language digits using convolutional neural network,'' *Proc. Comput. Sci.*, vol. 143, pp. 611–618, Jan. 2018.

[31] M. S. Islalm, M. M. Rahman, M. H. Rahman, M. Arifuzzaman, R. Sassi, and M. Aktaruzzaman, ''Recognition Bangla sign language using convolutional neural network,'' in *Proc. Int. Conf. Innov. Intell. Informat., Comput., Technol. (3ICT)*, Sep. 2019, pp. 1–6.

[32] S. A. Khan, A. D. Joy, S. M. Asaduzzaman, and M. Hossain, ''An efficient sign language translator device using convolutional neural network and customized ROI segmentation,'' in *Proc. 2nd Int. Conf. Commun. Eng. Technol. (ICCET)*, Apr. 2019, pp. 152–156.

[33] S. Nawshin, N. Saif, A. S. Mohammad, and M. Jameel, ''Protik: Bangla sign language teaching aid for children with impaired hearing,'' in *Proc. IEEE Region 10 Symp. (TENSYMP)*, Jun. 2020, pp. 440–443.

[34] F. S. Shishir, T. Hossain, and F. M. Shah, ''EsharaGAN: An approach to generate disentangle representation of sign language using InfoGAN,'' in *Proc. IEEE Region 10 Symp. (TENSYMP)*, Jun. 2020, pp. 1383–1386.

[35] M. M. Hasan, A. Y. Srizon, and M. A. M. Hasan, ''Classification of Bengali sign language characters by applying a novel deep convolutional neural network,'' in *Proc. IEEE Region 10 Symp. (TENSYMP)*, Jun. 2020, pp. 1303–1306.

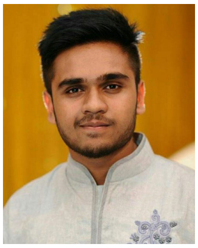
[36] M. A. Rahaman, M. Jasim, M. H. Ali, and M. Hasanuzzaman, ''Bangla language modeling algorithm for automatic recognition of hand-sign-spelled Bangla sign language,'' *Frontiers Comput. Sci.*, vol. 14, no. 3, pp. 1–20, Jun. 2020.

[37] M. Yeasin, S. K. Nath, M. I. H. Patwary, S. Hossain, S. Hossain, R. Ahmed, S. R. Ahmed, and M. H. Azad, ''Design and implementation of Bangla sign language translator,'' in *Proc. 5th Int. Conf. Adv. Electr. Eng. (ICAEE)*, Sep. 2019, pp. 815–820.

[38] M. A. Rahaman, M. Jasim, M. H. Ali, and M. Hasanuzzaman, ''Real-time computer vision-based Bengali sign language recognition,'' in *Proc. 17th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2014, pp. 192–197.

[39] S. M. Rayeed, G. W. Akram, S. T. Tuba, G. S. Zilani, H. Mahmud, and M. K. Hasan, ''Bangla sign digits recognition using depth information,'' in *Proc. 14th Int. Conf. Mach. Vis. (ICMV)*, Mar. 2022, p. 36.

[40] M. M. Hasan and Sk. M. M. Ahsan, "Bangla sign digits recognition using HOG feature based multi-class support vector machine," in *Proc. 4th Int. Conf. Electr. Inf. Commun. Technol. (EICT)*, Dec. 2019, pp. 1–5.
[41] M. A. Rahaman, M. Parvez Hossain, M. M. Rana, M. Arifur Rahman, and T. Akter, "A rule based system for Bangla voice and text to Bangla sign language interpretation," in *Proc. 2nd Int. Conf. Sustain. Technol. for Ind. 4.0 (STI)*, Dec. 2020, pp. 1–6.
[42] D. Tasmere, B. Ahmed, and M. M. Hasan, "Bangla sign digits: A dataset for real time hand gesture recognition," in *Proc. 11th Int. Conf. Electr. Comput. Eng. (ICECE)*, Dec. 2020, pp. 186–189.
[43] T. F. Ayshee, S. A. Raka, Q. R. Hasib, M. Hossain, and R. M. Rahman, "Fuzzy rule-based hand gesture recognition for Bengali characters," in *Proc. IEEE Int. Advance Comput. Conf. (IACC)*, Feb. 2014, pp. 484–489.
[44] S. T. Ahmed and M. A. H. Akhand, "Bangladeshi sign language recognition using fingertip position," in *Proc. Int. Conf. Med. Eng., Health Informat. Technol. (MediTec)*, Dec. 2016, pp. 1–5.
[45] J. Uddin, F. N. Arko, N. Tabassum, T. R. Trisha, and F. Ahmed, "Bangla sign language interpretation using bag of features and support vector machine," in *Proc. 3rd Int. Conf. Electr. Inf. Commun. Technol. (EICT)*, Dec. 2017, pp. 1–4.
[46] R. A. Nihal, S. Rahman, N. M. Broti, and S. Ahmed Deowan, "Bangla sign alphabet recognition with zero-shot and transfer learning," *Pattern Recognit. Lett.*, vol. 150, pp. 84–93, Oct. 2021.
[47] K. E. Aziz, A. Wadud, S. Sultana, M. A. Hussain, and A. Bhuiyan, "Bengali sign language recognition using dynamic skin calibration and geometric hashing," in *Proc. 6th Int. Conf. Informat., Electron. Vis. 7th Int. Symp. Comput. Med. Health Technol. (ICIEV-ISCMHT)*, Sep. 2017, pp. 1–5.
[48] S. S. Shanta, S. T. Anwar, and M. R. Kabir, "Bangla sign language detection using SIFT and CNN," in *Proc. 9th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2018, pp. 1–6.
[49] M. A. Hossen, A. Govindaiah, S. Sultana, and A. Bhuiyan, "Bengali sign language recognition using deep convolutional neural network," in *Proc. Joint 7th Int. Conf. Informat., Electron. Vis. (ICIEV) 2nd Int. Conf. Imag., Vis. Pattern Recognit. (icIVPR)*, Jun. 2018, pp. 369–373.
[50] A. J. Rony, K. H. Saikat, M. Tanzeem, and F. M. R. H. Robi, "An effective approach to communicate with the deaf and mute people by recognizing characters of one-hand Bangla sign language using convolutional neural-network," in *Proc. 4th Int. Conf. Electr. Eng. Inf. Commun. Technol. (iCEEiCT)*, Sep. 2018, pp. 74–79.
[51] A. M. Rafi, N. Nawal, N. S. N. Bayev, L. Nima, C. Shahnaz, and S. A. Fattah, "Image-based Bengali sign language alphabet recognition for deaf and dumb community," in *Proc. IEEE Global Humanitarian Technol. Conf. (GHTC)*, Oct. 2019, pp. 1–7.
[52] D. Tasmere and B. Ahmed, "Hand gesture recognition for Bangla sign language using deep convolution neural network," in *Proc. 2nd Int. Conf. Sustain. Technol. Ind. 4.0 (STI)*, Dec. 2020, pp. 1–5.
[53] N. Saquib and A. Rahman, "Application of machine learning techniques for real-time sign language detection using wearable sensors," in *Proc. 11th ACM Multimedia Syst. Conf.*, May 2020, pp. 178–189.
[54] A. Sayeed, M. M. Hasan, and A. Y. Srizon, "Bengali sign language characters recognition by utilizing transfer learned deep convolutional neural network," in *Proc. 11th Int. Conf. Electr. Comput. Eng. (ICECE)*, Dec. 2020, pp. 423–426.
[55] P. P. Urmee, M. A. A. Mashud, J. Akter, A. S. M. M. Jameel, and S. Islam, "Real-time Bangla sign language detection using xception model with augmented dataset," in *Proc. IEEE Int. WIE Conf. Electr. Comput. Eng. (WIECON-ECE)*, Nov. 2019, pp. 1–5.
[56] M. Hasan, T. H. Sajib, and M. Dey, "A machine learning based approach for the detection and recognition of Bangla sign language," in *Proc. Int. Conf. Med. Eng., Health Informat. Technol. (MediTec)*, Dec. 2016, pp. 1–5.
[57] A. M. J. Jim, I. Rafi, M. Z. Akon, and A. A. Nahid, "KU-BdSL: Khulna University Bengali sign language dataset," Mendeley Data, V4, 2023, doi: 10.17632/scpvm2nbkm.4.
[58] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[59] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1–9.
[60] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[61] P. Louridas and C. Ebert, "Machine learning," *IEEE Softw.*, vol. 33, no. 5, pp. 110–115, Sep. 2016.
[62] Y. Yu, M. Yao, J. Huang, and X. Xiao, "When process analysis technology meets transfer learning: A model transfer strategy between different spectrometers for quantitative analysis," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–19, 2024.
[63] Y. Yu and M. Yao, "A portable NIR system for nondestructive assessment of SSC and firmness of nanguo pears," *LWT*, vol. 167, Sep. 2022, Art. no. 113809.
[64] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.
[65] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," 2015, *arXiv:1511.07289*.
[66] P. Baldi and P. J. Sadowski, "Understanding dropout," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 1–9.
[67] 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. Knowledge Discovery Data Mining*, Aug. 2016, pp. 1135–1144.
[68] B. ChandraKarmokar, K. M. Rokibul Alam, and M. Kibria Siddiquee, "Bangladeshi sign language recognition employing neural network ensemble," *Int. J. Comput. Appl.*, vol. 58, no. 16, pp. 43–46, Nov. 2012.
[69] R. A. Nihal, N. M. Broti, S. A. Deowan, and S. Rahman, "Design and development of a humanoid robot for sign language interpretation," *Social Netw. Comput. Sci.*, vol. 2, no. 3, pp. 1–17, May 2021.
[70] A. A. J. Jim, I. Rafi, M. Z. Akon, U. Biswas, and A.-A. Nahid, "KU-BdSL: An open dataset for Bengali sign language recognition," *Data Brief*, vol. 51, Dec. 2023, Art. no. 109797.
[71] Y. Yu and M. Yao, "Is this pear sweeter than this apple? A universal SSC model for fruits with similar physicochemical properties," *Biosyst. Eng.*, vol. 226, pp. 116–131, Feb. 2023.



ABDULLAH AL JAID JIM received the B.Sc. (Eng.) degree in electronics and communication engineering from Khulna University, Khulna, Bangladesh, in 2022. He is a Lecturer with the Department of Electrical and Electronics Engineering, Trust University, Barishal, Bangladesh. He received the Merit Scholarship from Khulna University, from 2017 to 2022. He is the author of four articles and has made an image based Bengali sign language dataset. His research interests include robotics, the IoT, and machine learning. Integration of AI with everyday devices is one of his major goals.

IBRAHIM RAFI received the bachelor's degree in electronics and communication engineering from Khulna University, in 2023. He is committed to staying up-to-date with the latest technological advancements and making contributions to the ever-evolving fields of artificial intelligence and digital communication. His research interests include large language models, transformers, human activity recognition, and social media analysis.

JUN JIAT TIANG received the degree in electronics engineering from Multimedia University, Malaysia, the master's degree from the University of Science, Malaysia, and the Ph.D. degree from Universiti Kebangsaan Malaysia (UKM). He was a Design Automation Engineer with the Chipset Structural Design Team, Intel Microelectronics (M) Sdn. Bhd., Penang, Malaysia, from September 2004 to July 2005; and an Electronics Engineer with the Global Technology Development Group, Motorola Technology Sdn. Bhd., Penang, from May 2006 to May 2007. He has vast experience while working as the Project Leader in various research grants, such as TM Research and Development, from 2017 to 2020; Research Grant, Ministry of Science, Technology and Innovation (MOSTI), from 2008 to 2010; and Research Grant Mini Fund, in 2016. He is currently a Senior Lecturer and a Researcher with the Faculty of Engineering, Multimedia University. His research interests include RFID, microwave circuits, antenna, and propagation. He received the Gold Medal at the 23rd International Invention, Innovation, and Technology Exhibition (ITEX) 2012, Kuala Lumpur, Malaysia, in May 2012; and the Silver Medal at the Malaysian Technology Expo (MTE) 2013, Kuala Lumpur, in February 2013.

UZZAL BISWAS received the Bachelor of Science and Master of Science degrees (Hons.) in electronics and communication engineering from Khulna University, Bangladesh, in 2011 and 2014, respectively, and the Doctor of Philosophy degree in biomedical engineering from the University of New South Wales (UNSW), Sydney, Australia, in 2021. In 2011, he joined in Electronics and Communication Engineering (ECE) Discipline, as a Faculty Member. Currently, he is a Professor with the ECE Discipline, Khulna University, Khulna, Bangladesh. In 2016, he received the prestigious Endeavour Postgraduate Research Scholarship from the Education Ministry of Australian Government to pursue the Ph.D. degree from UNSW. His research interests include biomedical signal processing, telehealth-based remote patient monitoring, data mining, time series analysis, machine learning-based classification, and prediction modeling. In 2015, he received the President Gold Medal Award from Khulna University, for securing first position with distinction and obtaining the highest CGPA in the undergraduate program among all disciplines under the Science, Engineering and Technology (SET) School of Khulna University.

ABDULLAH-AL NAHID received the B.Sc. degree in electronics and communication engineering from Khulna University, Khulna, Bangladesh, in 2007, the M.Sc. degree in telecommunication engineering from the Institute for Telecommunication Research, University of South Australia, Mawson Lakes, SA, Australia, in 2012, and the Ph.D. degree in machine learning from Macquarie University, Sydney, NSW, Australia, in 2018.

He is currently with the Electronics and Communication Engineering Discipline, Khulna University. His research interests include machine learning, biomedical image processing, data classification, and smart grids.