Research Article

Soft Voting-based Ensemble Model for Bengali Sign Gesture Recognition

Md Abdur Rahim[1], Jungpil Shin[2] and Keun Soo Yun[3,*]

[1]Pabna University of Science and Technology, Bangladesh
rahim@pust.ac.bd
[2]The University of Aizu, Aizuwakamatsu, Japan
jpshin@u-aizu.ac.jp
[3]Ulsan College, South Korea.
ksyun@uc.ac.kr
*Correspondence: ksyun@uc.ac.kr

Received: 13th February 2022; Accepted: 25th March 2022; Published: 1st April 2022

Abstract: Human hand gestures are becoming one of the most important, intuitive, and essential means of recognizing sign language. Sign language is used to convey different meanings through visual-manual methods. Hand gestures help the hearing impaired to communicate. Nevertheless, it is very difficult to achieve a high recognition rate of hand gestures due to the environment and physical anatomy of human beings such as light condition, hand size, position, and uncontrolled environment. Moreover, the recognition of appropriate gestures is currently considered a major challenge. In this context, this paper proposes a probabilistic soft voting-based ensemble model to recognize Bengali sign gestures. We have divided this study into pre-processing, data augmentation and ensemble model-based voting process, and classification for gesture recognition. The purpose of pre-processing is to remove noise from input images, resize it, and segment hand gestures. Data augmentation is applied to create a larger database for in-depth model training. Finally, the ensemble model consists of a support vector machine (SVM), random forest (RF), and convolution neural network (CNN) is used to train and classify gestures. Whereas, the ReLu activation function is used in CNN to solve neuron death problems and to accelerate RF classification through principal component analysis (PCA). A Bengali Sign Number Dataset named "BSN-Dataset" is proposed for model performance. The proposed technique enhances sign gesture recognition capabilities by utilizing segmentation, augmentation, and soft-voting classifiers which have obtained an average of 99.50% greater performance than CNN, RF, and SVM individually, as well as significantly more accuracy than existing systems.

Keywords: *Augmentation; Convolutional Neural Network (CNN); Deep ensemble; Hand gesture; Sign Language; Support Vector Machine (SVM); Random Forest (RF)*

---

1. Introduction

The knowledge of sign language is essential for expressing importance, awareness, and sensitivity towards deaf culture and hard-of-hearing (HoH) people. HoH refers to people who can hear with a hearing aid and whose hearing is mild to more severe, whereas deaf people rarely hear or do not hear at all. Hearing loss affects communication, however, sign language develops a strong perception of deaf culture and increases language comprehension and acceptance among others. Worldwide, 466 million people suffer from hearing loss, including 34 million children [1]. However, sign language is not a universal language; it is used differently in different nations or locations. It is a complex medium of communication that includes hand gestures, body parts, and facial expressions that give the deaf community the ability to communicate thoughts and feelings effectively. American Sign Language (ASL), Pidgin Signed English (PSE), and Bengali Sign Language (BSL) are three of the most common varieties of sign language. Despite its importance,

Md Abdur Rahim, Jungpil Shin and Keun Soo Yun, "Soft Voting-based Ensemble Model for Bengali Sign Gesture Recognition", *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN: 2516-0281, Online ISSN: 2516-029X, pp. 41-49, Vol. 6, No. 2, 1st April 2022, Published by International Association for Educators and Researchers (IAER), DOI: 10.33166/AETiC.2022.02.003, Available: http://aetic.theiaer.org/archive/v6/v6n2/p3.html.

Bengali sign language was not as recognized as other popular languages. However, Bengali is one of the most widely spoken languages in the world and the second-most in the Indian subcontinent. About 228 million people speak Bengali as their mother tongue, 37 million people speak their second language [2]. People sacrificed their lives for the Bengali language on 21st February 1952 and after the independence of the People's Republic of Bangladesh, UNESCO declared February 21st to be International Mother Language Day in 1999 [3]. Researchers have been interested in the recognition of Bengali sign language since it is the world's fifth most spoken, popular, and valued language.

The Bengali language has a total of 50 characters comprising of eleven vowels and thirty-nine consonants. There are ten modifiers, ten numbers, and around 300 compound characters in it [4]. The recognition of Bengali sign numbers are the emphasis of this article. For the recognition of Bengali numerals' sign gestures, we presented a deep ensemble-based model. The ensemble learning method is used to reduce the likelihood of the model choosing the wrong class. However, it consists of multiple classifiers which are grouped according to a specific method for solving a specific problem. If a weak classifier makes a mistake, additional weak classifiers can rectify it. Moreover, this ensemble model integrates deep and machine learning techniques such as CNN, RF, and SVM. The main goal of this research is to accurately recognize the gestures of Bengali signs. The following are the contributions of this research work:

1. Firstly, a Bengali Sign Number Dataset named "BSN-Dataset" has been created which contains hand gestures of 10 Bengali numbers in different light and complex backgrounds.
2. Secondly, an efficient segmentation approach i.e., YCbCr is applied to identify gesture signs. It is considered to be the most efficient for dealing with the challenges of vision-oriented systems such as gesture detection, complex environments, and changing lighting conditions.
3. Thirdly, data augmentation techniques such as rotation and transformation have been applied to enlarge the dataset images for training purposes. This helps to solve the problem of class imbalances in the classification and prevents data deficiencies for better models.
4. Finally, the soft voting-based ensemble model offers for gesture recognition. This model consists of three different methods such as CNN, RF, and SVM. However, the combined classification results of each method predict output based on a majority vote. In addition, a comparison is made between our results and similar work of other researchers.

This paper is presented as follows. Section 1 describes the rationale and purpose of this study. A review of the literature related to this study is explained in Section 2. Section 3 discusses the dataset, the augmentation process, and the description of the proposed ensemble model. The results and evaluation of the proposed ensemble model with appropriate figures and tables are mentioned in Section 4, and finally the summary of this study is given in the conclusion section.

2. Related Work

A brief discussion of sign language recognition in this section is divided into two ways, review on Bengali Sign Language and review on other sign languages.

Based on isolated Bangla characters, a hand gesture dataset [5] was proposed. There are 50 sets of 36 Bangla characters in the dataset, each comprising six vowels and thirty consonants. In both training and validation, the authors achieved 92.65% and 94.74% accuracy. The authors proposed a Bengali language modelling to recognize hand-spelled Bengali signs in [6]. The model obtained 93.50%, 95.50%, and 90.50% accuracy in Bengali written word, compound number and sentence recognition gestures respectively. However, the desire for recognition can be improved by introducing deep learning methods. In [7], a transfer and Zero-shot learning technique for Bengali sign language recognition was presented. Using the linear differential analysis method, the authors evaluated 35,149 images and achieved an overall accuracy of 93.68%. In [8], the authors developed a real-time Bengali sign language translator that incorporated the translation of ordinarily spoken Bengali into Bengali signs. Overall, the accuracy of recognition is one of the key issues of sign language recognition. The performance of a convolutional neural network in recognizing numbers and alphabets independently, as well as blending them, was assessed using 10-fold cross-validation on the created dataset [9]. The small number of images in the collection, however, may have an impact on recognition accuracy. Deep learning models were used by Podder *et al.* [10] to classify BdSL

Alphabets and numerals. The authors reached a high level of accuracy of 99.99%, and this dataset is freely available to researchers to assist and stimulate further study. To create and train systems for identifying Bangladeshi sign language, a quicker R-CNN model was proposed in [11]. The identification accuracy of sign alphabets was 98.2%, however, the pattern had numerous similarities.

In addition, many researchers are conducting research in a variety of sign languages. Wadhawan *et al.* suggested a strong models for static signs recognition in [12], and the authors achieved a training accuracy of 99.72%. In [13], emphasis is placed on the relevance of artificial intelligence (AI) technology in sign language. The author goes on sign language capturing and recognition, translation, presentation, and a full assessment of current approaches. Tolentino *et al.* [14] have created a skin-color modelling-based sign language learning aid for beginners. The authors were able to recognize the ASL alphabet with 90.04%. A hybrid segmentation-based sign word recognition system was developed in [15]. YCbCr and SkinMask were proposed and achieved 97.28% accuracy using the Support Vector Machine (SVM) classifier. The Otsu segmentation and CNN approaches were used to create ASL alphabet recognition and sentence interpretation in [16]. Due to the small number of dataset images, the accuracy of the recognition may be impaired. The CNN-based ASL alphabet recognition interface was developed in [17]. The accuracy of recognition may be compromised as a result of image overfitting during training. In this paper, to achieve the best accuracy, we present an ensemble model that combines both machine and deep learning to recognize hand sign gestures.

3. Proposed Methodology

This section describes the proposed method of Bengali sign number recognition (BSNR). We have followed the steps of pre-processing, image augmentation, trained ensemble models and voting process for model evaluation to recognize the Bengali sign number. Figure 1 shows the accuracy of the recognition of the BSNR system.

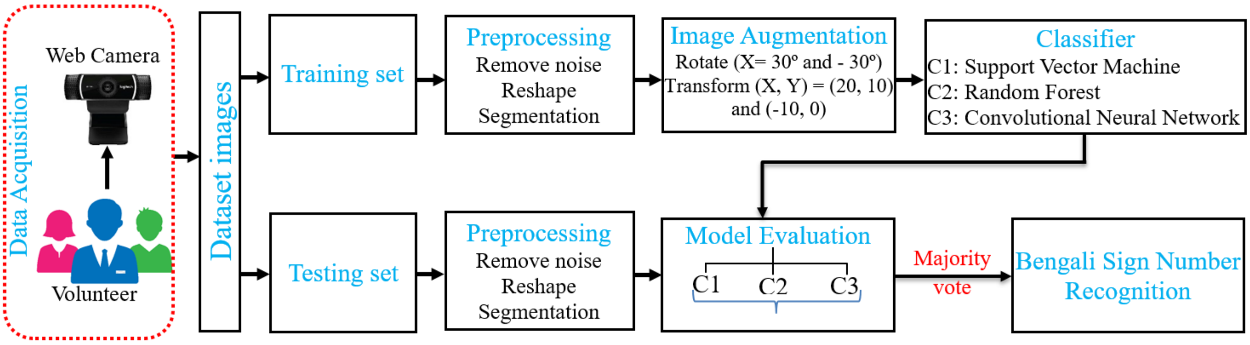

Figure 1. Basic block diagram of Bengali sign numbers recognition system

3.1. Dataset Description

In this study, we have created a sign gesture dataset of Bengali numbers. Figure 2 presents the hand sign gesture of Bengali numbers. The Bengali sign gesture dataset consists of 10 (ten) category and contains 500 images from each person for each category which are captured by changing the position and size of the hand within the image frame. The colour image was taken and the size of the image is 200 x 200 pixels. Sign gestures are performed by 10 (ten) volunteers in different lighting and complex backgrounds. Both male (8) and female (2) volunteers participated and ranged in age from 20 to 40 years. Each class contains 5000 images with a total dataset of 50,000 images. For evaluation and model performance, the 70% dataset images were used for training and the remaining 30% to 15% for validation and 15% for testing.

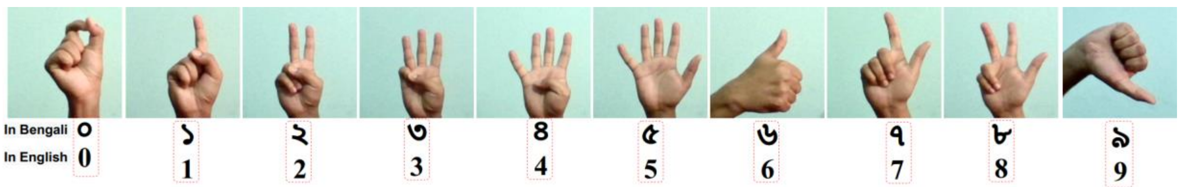

Figure 2. Examples of Bengali numbers (0 to 9 in English) through hand gestures

3.2. Pre-processing

The Gaussian blurring technique is applied to remove noise from the input image. Therefore, the input images were resized to 50x50 pixels for segmentation, augmentation and fed for the classification models. However, the segmentation of gesture parts from input images is essential for sign language recognition. Thus, in this paper, we applied YCbCr techniques to segment sign gesture images. The input images were converted to grayscale images in the YCbCr color space. YCbCr contains elements of luminance, blue and red which are defined as Y, Cb, and Cr respectively. The Cr component is used for subsequent pre-processing. In this case, the Cr images were converted to binary images. To convert the grayscale image into a binary image, the threshold value was set to 128 to redistribute the pixel values from 0-127 to 0 and 128-255 to 255. However, the grayscale value ranges from 0 to 255, where 0 is defined as black and 255 as white. Also, pixel values are defined as black for the range 0 to 127 and white for the range 128 to 255. Finally, we obtained the segmented image by applying the erosion that removes the boundary of the foreground pixels.

3.3. Image Augmentation

The performance of most machine and deep learning models depends on the quantity and variety of data [18]. In this case, data augmentation methods have become essential to reduce dependency on data preparation and creation and to create more accurate machine learning models. Thus, we have added more data for training models to improve the accuracy of model predictions. Augmentation techniques such as rotation and transforms are applied to generate more dataset images to prevent data deficits for the proposed models. To ensure the variation in orientation, we employ rotation that rotates the image in 30º and -30º directions, and performs the transformation by shifting certain amounts of pixels, such as (20, 10) and (-10, 0) in X and Y directions. However, two rotation and two transforms techniques are applied for each input image, therefore, each class contains a total of 17,500 images where 3500 images are segmented and 14,000 augmented images. Table 1 shows a description of the type of data augmentation.

Table 1. Description of different types of data augmentation techniques

Augmentation type	Description	Experimental Values
Rotate Xº	Rotate the image by a certain amount	X=30º and -30º
Transforms (X, Y)	Shift a certain amount of pixel values in the X and Y directions of the input image	X, Y= (20, 10) and (-10, 0)

3.4. Ensemble Model

The ensemble method is a technique that combines different methods to create the best predictive model. An ensemble can predict better than any single model. It reduces the dispersion of model performances. In this paper, we proposed a model based on deep ensemble learning for BSNR. We used three classifiers to learn the ensemble and these classifiers are transformed into a possibility for soft voting. However, recognition of Bengali sign numbers may be erroneous due to finger angle, change of light, handshape, and uncontrolled environment. To solve these problems, we have selected three classification models such as RF, SVM, and CNN. However, these three classifiers can complement each other in BSLR.

3.4.1. Convolutional Neural Network (CNN)

A convolutional neural network (CNN) is often used for recognizing images or gestures, performing the generative and descriptive tasks through one or more convolutional levels [19]. The proposed CNN architecture consists of five convolution and max-pooling layers as well as one fully connected layer. There are 32 filters for the first convolutional layer and 64 filters for the rest. It has kernel size 3 and padding 1 and here ReLU is employed as the activation function. Also, there is a 20% dropout after max pooling for different layers. The output is flattened into an array and transmitted to a fully connected layer. It consists of flattened, dense, and dropout layers. After flattening the data, it reaches the dense layer that drives the classification process. The dropout layer is used to prevent overfitting by randomly ignoring neurons. The output of this layer uses ReLU following batch normalization. The design of the proposed CNN model is shown in Figure 3.

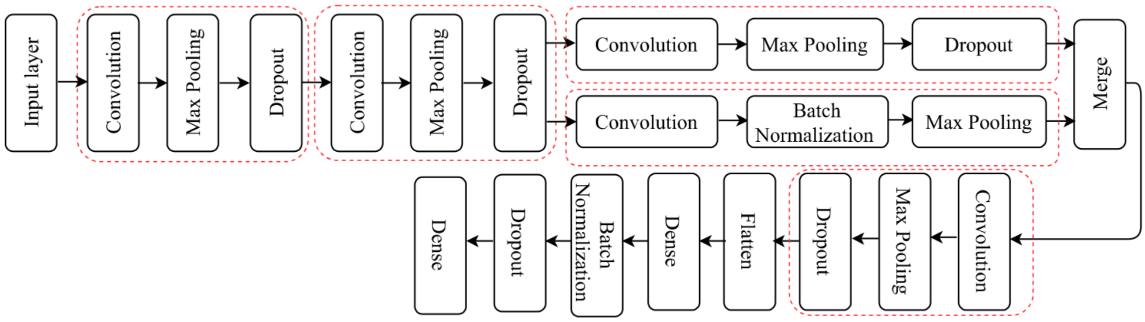

Figure 3. The Architecture of the proposed CNN model

3.4.2. Random Forest (RF)

In a random forest, ensemble learning is utilized to handle regression and classification problems [20]. It combines decision trees and establishes results based on those trees. The results of the predictions are calculated by taking the average or the mean. However, the use of many decision-making trees slows down the RF results and very few decision-making trees lose their accuracy. In this work, image data levels need to be reduced for RF classification. Thus, we have used principal component analysis (PCA) which reduces the data level of the image. We implemented Bootstrap to divide training data that effectively prevent over-fitting. However, the training dataset contains observations and features that are used to make predictions. Finally, all predicted results of the decision tree algorithm were consolidated with a vote, where the maximum would be selected as the final output.

3.4.3. Support Vector Machine (SVM)

The Support Vector Machine focuses on n-dimensional space data points with a specific coordinate value for each feature [21]. This makes a good classification by finding the hyperplane to indicate the distance between the two types of samples closest to the plane on both sides. It uses kernel techniques to transform data. After transformation, it determines an optimal boundary between the possible outputs. In essence, the kernel strategy transforms data to a higher dimension through effective and less expensive ways. The radial basis kernel (RBF) function is used in this study. Its notable feature is that it stores support vectors during training and not in the case of entire datasets. The functionality of the RBF kernel in SVM can result in localization and limited response over the entire range of the main axis. The kernel function is defined in Equation (1) as

$$K(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right) \quad (1)$$

Where RBF kernel K in two samples $x \ and \ x'$ represents as feature vectors, $\|x - x'\|$ is defined as the squared Euclidean distance between these feature vectors, and σ as free parameters.

3.5. Soft Voting Classifier

A voting classifier predicts a result based on the class having the highest possibility of being output. It is based on an ensemble technique, which incorporates the results of each voting classifier classification and predicts the output based on the majority vote [22]. In this paper, we have introduced a soft voting technique that predicts the outcome using the average of the probabilities given in that category. For this, a weighted soft voting is applied to determine the results of the classification. The results of the weighted soft voting are expressed in Equation (2) as follows,

$$H(P) = V_{max} \sum_{i=1}^{N} w_i \times h_i(P);\ w_i \geq 0 \ and \ \sum_{i=1}^{N} w_i = 1 \quad (2)$$

Where the highest value obtains from V_{max} and the prediction probability of the classification is $h_i(P)$. The output data, weight, number, and level of classifiers are P, w_i, N, and i, respectively.

4. Results and Evaluation

This section describes and illustrates the experimental settings, pre-processing and segmentation, data augmentation, analysis of individual and ensemble architectures, performance analysis and evaluation, and comparisons with previous work.

4.1. Pre-processing and Augmentation Process

For machine learning and deep network processing, we applied input images that could detect significant features. First, the noise in the input image was removed using the Gaussian blurring technique. Therefore, the input images were resized as 50x50 for the process of segmentation, augmentation, and classification of single and ensemble methods. Here, we applied the YCbCr technique to segment hand gestures from noise-free images. Figure 4 illustrates an example of a segmented image. Therefore, various augmentation techniques are applied to increase the size of the dataset for the purpose of training the proposed models. Moreover, the testing dataset images are used for model evaluation.

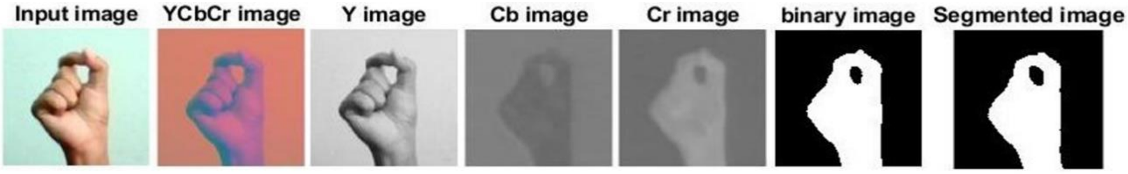

Figure 4. An example of hand gesture segmentation of an input image

To improve performance, we have enlarged our datasets to increase the generalizability of models by applying augmentation techniques like rotation and transforms. BSN-dataset images have been divided into 70% and 30% for training and testing purposes. Therefore, a total of 35,000 images are employed for training. In addition, to create data variability, 140,000 augmented images have been added whereas the total dataset images for training is 175,000 which plays an effective role to improve model accuracy. The trained models are then used to classify the sign numbers through the voting classifier. Figure 5 shows an example of the augmentation of an input image.

Figure 5. Examples of different augmentation of an image (left to right: segmented image, rotate 300, rotate -300, transform (20, 10), and transform (-10, 0)

4.2. Results Analysis of Ensemble Model

The purpose of this study is to accurately predict the Bengali sign number through a deep ensemble model. The performance of this model has been evaluated using the train, test, and validation sets. The train set is for model training and the validation set is for performance testing. The test set is used to test the final performance of the trained model. However, the key classification metrics for evaluating the performance of the proposed method are accuracy, recall, precision, and F1-score. The proposed model was trained and validated with augmented datasets for 30 epochs. In the beginning, both the loss of validation and the loss of the train were high and gradually decreased as the training process continued. However, the increase in validation accuracy and the accuracy of the train is proportional to the increase in the epochs. The highest accuracy was 99.97% of validity and the accuracy of training was 99.50% accuracy. The confusion matrix of the ensemble model is shown in Figure 6. Table 2 presents the accuracy of the predictions of the proposed model and the models are described individually.

We observed that the results of our proposed ensemble model achieved the recognition accuracy of Bengali sign numbers is 99.50%. The results of the proposed and individual models are described in detail in Table 2. The best performance accuracy of the CNN model is 99.44%, whereas SVM is 97.39% and RF is 95.56%. Therefore, the CNN model has performed as a major classifier in the voting ensemble model where other classifiers play a supporting role. The accuracy of the maximum recognition is 100% of the numbers 5, this may be due to less confusion between the numbers. Table 3 presents the accuracy of the gestures of each Bengali number sign and it includes the accuracy, precision, and f1-score. However, the average accuracy of the precision, recall, and F1-score are 99.55%, 99.35%, and 99.45% respectively.

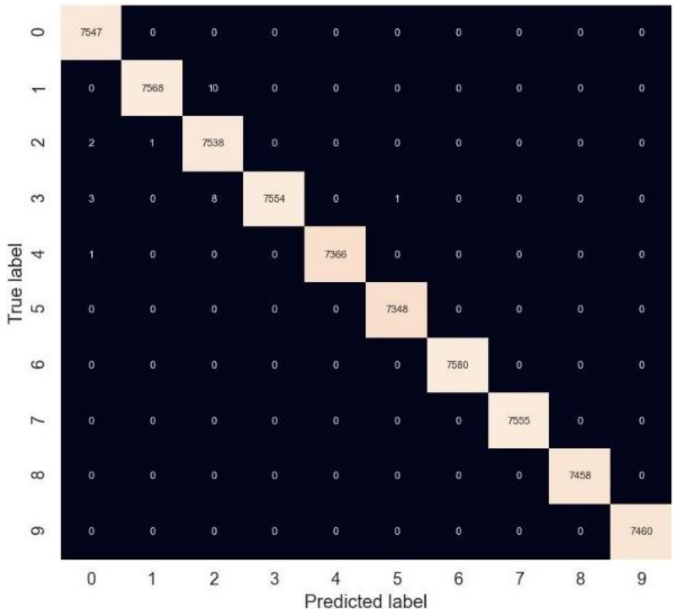

Figure 6. The confusion matrix of the ensemble model

Table 2. Descriptions and equation of classification metrics

Methods	Accuracy	Precision	Recall	F1-Score
SVM	96.39%	96.12%	96.08%	95.94%
RF	95.56%	95.32%	95.27%	94.91%
CNN	99.44%	99.79%	99.65%	99.72%
Ensemble model	99.50%	99.80%	99.69%	99.75%

Table 3. Classification performances of each gesture based on precision, recall, and F1-score

Gestures	Precision	Recall	F1-score
০	99.32	99.86	99.59
১	99.60	99.66	99.63
২	99.87	99.67	99.77
৩	99.87	99.60	99.73
৪	99.31	97.96	98.63
৫	100	100	100
৬	99.32	99.93	99.62
৭	99.73	99.80	99.76
৮	99.87	99.80	99.83
৯	98.61	97.26	97.93
Average	99.55	99.35	99.45

4.3. Comparison with Previous Work

We compared the efficiency of the proposed ensemble method with the dataset [9] and state-of-the-art methods. In [9], the author created a Bangladeshi sign language (BdSL) dataset with basic characters and numbers containing 23864 and 7052 images, respectively. In this paper, we considered only Bengali numbers for comparison. To implement the proposed model, we applied pre-processing methods, segmentation, and augmentation techniques that extend the dataset from 7052 to 26804 where 24690 for training. Table 4 presents the performance accuracy of the proposed model using proposed and benchmark datasets. We observed that the proposed ensemble method achieved 99.50% accuracy using our dataset and 100% accuracy using Bangladeshi Sign Language (BdSL) dataset. However, the proposed model shows high precision performance in determining accuracy. A comparative analysis between the performance of the various sign alphabets and numbers recognition models reported in related work and our proposed model is shown in Table 5.

Table 4. Training and testing accuracy of proposed and BdSL datasets [9]

Dataset	Total no. of Images		Recognition Accuracy (%)
	Training	Testing	
Proposed (Bengali Sign numbers)	175,000	7500	99.50%
BdSL Number [9]	24690	1057	100%

Table 5. Comparison of the performance accuracy of our proposed model with state-of-the-art methods

Reference	Gestures	Methods	Dataset size (Images)	Recognition accuracy (%)
[5]	BdSL	CNN	1800	92.65
[7]	BdSL	Transfer learning	35149	93.68
[10]	Bengali Sign Number	CNN	1075	95.00
[9]	Bengali Sign Number	Using our proposed model	26804	100
Proposed BSLR	Bengali Sign Number	Proposed Ensemble model	190000	99.50

5. Conclusion

This paper recognizes the Bengali sign numbers based on the soft voting-based ensemble model. It works on a voting system that combines the three classifiers SVM, RF, and CNN and classifies sign gestures by measuring majority votes. A dataset called BSN-dataset was created where 10 participants were given 10 tasks and 500 images were taken for each task. Therefore, dataset images are used for experimental and performance evaluation following the steps of pre-processing, segmentation, data augmentation and classification. The noise is removed from the input image and then resized. The YCbCr technique was then applied to segment the hand sign gesture. Subsequently, the dataset was augmented for the training of the deep ensemble model. The augmentation technique consists of 30º and -30º rotation and transformation of (X, Y) direction as (20, 10) and (-10, 0) respectively. From the results, we can say that the proposed method has achieved a very high accuracy of 99.50% compared to SVM (96.39%), RF (95.56%), CNN (99.44%), and state-of-the-art methods performed separately. The future goal is to use BSLR system to overcome the communication gap between the deaf population and the general public by utilizing the most accessible and adaptable technology, such as smartphones.

Acknowledgment

This work has been supported by the 2021 Research Fund of Ulsan College.

References

[1] Ning Ding, Sangsin Lee, Matan Lieber-Kotz, Jie Yang and Xue Gao, "Advances in genome editing for genetic hearing loss", *Advanced drug delivery reviews*, Vol. 168, pp. 118-133, 01 January 2021, Published by Elsevier, DOI: 10.1016/j.addr.2020.05.001.

[2] Abu Sayeed, Jungpil Shin, Md. Al Mehedi Hasan and Azmain Yakin Srizon, "BengaliNet: A Low-Cost Novel Convolutional Neural Network for Bengali Handwritten Characters Recognition", *Applied Sciences*, Vol. 11, No. 15, pp. 6845, 25 July 2021, Published by MDPI Multidisciplinary Digital Publishing Institute, DOI: 10.3390/app11156845, Available: https://www.mdpi.com/2076-3417/11/15/6845/htm.

[3] Anita Pandey, "Using mother tongues as building blocks in childhood education", *Childhood Education*, Vol. 90, No. 1, pp. 61-67, 23 December 2013, Published by Springer, DOI: 10.1080/00094056.2014.872517.

[4] Ritesh Sarkhel, Nibaran Das, Amit K. Saha and Mita Nasipuri, "A multi-objective approach towards cost effective isolated handwritten Bangla character and digit recognition", *Pattern Recognition*, Vol. 58, pp. 172-189, 1st October 2016, Published by Elsevier, DOI: 10.1016/j.patcog.2016.04.010.

[5] Md Sanzidul Islam, Sadia Sultana Sharmin Mousumi, Nazmul A. Jessan, AKM Shahariar Azad Rabby and Sayed Akhter Hossain, "Ishara-Lipi: The first complete multipurpose open access dataset of isolated characters for Bangla sign language", In *Proceedings of the 2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, 03 December 2018, Electronic ISBN: 978-1-5386-8207-4, Print on Demand (PoD) ISBN: 978-1-5386-8208-1, DOI: 10.1109/ICBSLP.2018.8554466, pp. 1-4, Published by IEEE.

[6] Muhammad Aminur Rahaman, Mahmood Jasim, Md Ali and Md Hasanuzzaman, "Bangla language modeling algorithm for automatic recognition of hand-sign-spelled Bangla sign language", *Frontiers of Computer Science*, Vol. 14, No. 3, pp. 1-20, 07 December 2019, Published by Springer Science and Business Media, DOI. 10.1007/s11704-018-7253-3.

[7] Ragib Amin Nihal, Sejuti Rahman, Nawara Mahmood Broti and Shamim Ahmed Deowan, "Bangla Sign alphabet recognition with zero-shot and transfer learning", *Pattern Recognition Letters*, Vol. 150, pp. 84-93, October 2021, Published by Elsevier, DOI: 10.1016/j.patrec.2021.06.020.

[8] Md Tazimul Hoque, Md Rifat-Ut-Tauwab, Md Fasihul Kabir, Farhana Sarker, Mohammad Nurul Huda *et al.*, "Automated Bangla sign language translation system: Prospects, limitations and applications." In *Proceedings of the 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, 13-14 May 2016, Dhaka, Bangladesh, Electronic ISBN:978-1-5090-1269-5, Print on Demand(PoD) ISBN:978-1-5090-1270-1, DOI: 10.1109/ICIEV.2016.7760123, pp. 856-862, Published by IEEE.

[9] Md Shafiqul Islalm, Md Moklesur Rahman, Md Hafizur Rahman, Md Arifuzzaman, Roberto Sassi *et al.*, "Recognition Bangla sign language using convolutional neural network", In *Proceedings of the 2019 international conference on innovation and intelligence for informatics, computing, and technologies (3ICT)*, pp. 1-6, 2-23 September 2019, Electronic ISBN:978-1-7281-3012-5, Print on Demand (PoD) ISBN:978-1-7281-3013-2, Published by IEEE, DOI: 10.1109/3ICT.2019.8910301.

[10] Sanzidul Islam, Sadia Sultana Sharmin Mousumi, AKM Shahariar Azad Rabby, Sayed Akhter Hossain and Sheikh Abujar, "A potent model to recognize Bangla sign language digits using convolutional neural network", *Procedia Computer Science*, Vol. 143, pp. 611-618, 2018, Published by Elsevier BV, DOI: 10.1016/j.procs.2018.10.438.

[11] Oishee Bintey Hoque, Mohammad Imrul Jubair, Md Saiful Islam, Al-Farabi Akash and Alvin Sachie Paulson, "Real time Bangladeshi sign language detection using faster r-cnn", In *Proceedings of the 2018 international conference on innovation in engineering and technology (ICIET)*, pp. 1-6, 27-28 December 2018, Electronic ISBN:978-1-5386-5229-9, Print on Demand(PoD) ISBN: 978-1-5386-5230-5, Published by IEEE, DOI: 10.1109/CIET.2018.8660780.

[12] Ankita Wadhawan and Parteek Kumar, "Deep learning-based sign language recognition system for static signs", *Neural computing and applications, Vol.* 32, No. 12, pp. 7957-7968, 01 January 2020, Published by Springer, DOI: 10.1007/s00521-019-04691-y.

[13] Ilias Papastratis, Christos Chatzikonstantinou, Dimitrios Konstantinidis, Kosmas Dimitropoulos and Petros Daras, "Artificial intelligence technologies for sign language", *Sensors*, Vol. 21, No. 17, pp. 5843, 30 August 2021, Published by MDPI, DOI: 10.3390/s21175843, Available: https://www.mdpi.com/1424-8220/21/17/5843/htm.

[14] Lean Karlo S. Tolentino, Ronnie O. Serfa Juan, August C. Thio-ac, Maria Abigail B. Pamahoy, Joni Rose R. Forteza, *et al*. "Static sign language recognition using deep learning", *International Journal of Machine Learning and Computing*, Vol. 9, No. 6, pp. 821-827, 2019, Available: http://www.ijmlc.org/vol9/879-L0320.pdf.

[15] Md Abdur Rahim, Md Rashedul Islam and Jungpil Shin, "Non-touch sign word recognition based on dynamic hand gesture using hybrid segmentation and CNN feature fusion", *Applied Sciences*, Vol. 9, No. 18, pp. 3790, 10 December 2019, DOI: 10.3390/app9183790, Published by MDPI Multidisciplinary Digital Publishing Institute, Available: https://www.mdpi.com/2076-3417/9/18/3790/htm.

[16] Md Abdur Rahim, Jungpil Shin and Keun Soo Yun, "Hand Gesture-based Sign Alphabet Recognition and Sentence Interpretation using a Convolutional Neural Network", *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN: 2516-0281, Online ISSN: 2516-029X, pp. 20-27, Vol. 4, No. 4, 1st October 2020, Published by International Association for Educators and Researchers (IAER), DOI: 10.33166/AETiC.2020.04.003, Available: http://aetic.theiaer.org/archive/v4/v4n4/p3.html.

[17] Ahmed Kasapbaşi, Ahmed Eltaye Ahmed Elbushra, Omar Al-Hardanee and Arif Yilmaz, "DeepASLR: A CNN based Human Computer Interface for American Sign Language Recognition for Hearing-Impaired Individuals", *Computer Methods and Programs in Biomedicine Update*, Vol. 2, pp. 100048, 2022, DOI: 10.1016/j.cmpbup.2021.100048, Available: https://www.sciencedirect.com/science/article/pii/S2666990021000471.

[18] Connor Shorten and Taghi M. Khoshgoftaar, "A survey on image data augmentation for deep learning", *Journal of Big Data*, Vol. 6, No. 1, pp. 1-48, 6 July 2019, Published by Springer Open, DOI: 10.1186/s40537-019-0197-0, Available: https://link.springer.com/article/10.1186/s40537-019-0197-0.

[19] Juan C. Nunez, Raul Cabido, Juan J. Pantrigo, Antonio S. Montemayor and Jose F. Velez, "Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition", *Pattern Recognition*, Vol. 76, pp. 80-94, 1st April 2018, Published by Elsevier, DOI: 10.1016/j.patcog.2017.10.033.

[20] Junshi Xia, Pedram Ghamisi, Naoto Yokoya and Akira Iwasaki, "Random forest ensembles and extended multiextinction profiles for hyperspectral image classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 56, No. 1, pp. 202-216, 19 September 2017, Published by Institute of Electrical and Electronics Engineers Inc. DOI: 10.1109/TGRS.2017.2744662.

[21] Carl Gold and Peter Sollich, "Model selection for support vector machine classification", *Neurocomputing*, Vol. 55, No. 1-2, pp. 221-249, 1 September 2003, Published by Elsevier, DOI: 10.1016/S0925-2312(03)00375-8.

[22] Saloni Kumari, Deepika Kumar and Mamta Mittal, "An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier", *International Journal of Cognitive Computing in Engineering*, Vol. 2, pp. 40-46, June 2021, Published by Elsevier, DOI: 10.1016/j.ijcce.2021.01.001.