Undergraduate Conference on Intelligent Computing and Systems (UCICS 2025)
26-27 February, 2025; Varendra University, Rajshahi, Bangladesh

Hybrid Model-based Bangla Sign Language Recognition Using Machine Learning and Deep Transfer Learning Techniques

Md Anwarul Islam Himel
Dept. of CSE
Pabna University of Science and Technology Pabna, Bangladesh.
Email: himelhim31@gmail.com

Md Bayazid Hossain
Dept. of CSE
Pabna University of Science and Technology Pabna, Bangladesh.
Email: bayazid204@gmail.com

Md Abdur Rahim
Dept. of CSE
Pabna University of Science and Technology Pabna, Bangladesh.
Email: rahim@pust.ac.bd

Shabbir Mahmood
Dept. of CSE
Pabna University of Science and Technology Pabna, Bangladesh.
Email: shabbir.cse@pust.ac.bd

Abu Saleh Musa Miah
Dept. of CSE
The University of Aizu
Aizuwakamatsu, Japan
Email: abusalehcse.ru@gmail.com

Md. Imran Hossain
Dept. of ICE
Pabna University of Science and Technology Pabna, Bangladesh.
Email: imran05ice@pust.ac.bd

***Abstract*—Sign language serves as a critical communication medium for deaf and nonverbal individuals; however, its limited use by hearing individuals creates a significant communication gap. This study introduces a hybrid model for automated recognition of Bangla Sign Language (BdSL) to tackle this challenge. The proposed approach integrates an efficient gesture classification system using a Support Vector Machine (SVM) classifier and a convolutional neural network (CNN) leveraging deep transfer learning for feature extraction. By combining the strengths of traditional machine learning with deep learning, especially in scenarios with limited training data, the hybrid model achieves enhanced accuracy. When tested on the KU-BdSL dataset, the model demonstrated an exceptional accuracy of 99.7%, underscoring its effectiveness. This research highlights the potential of hybrid models in advancing sign language recognition and facilitate interaction among hearing and deaf communities.**

***Index Terms*—Bangla Sign Language, Machine Learning, Deep Learning, Data Augmentation, Transfer Learning**

I. Introduction

For people with hearing and speech impairments, sign language is an essential communication method. In contrast to spoken languages, which depend on audible sounds and written symbols, sign languages utilize visual gestures, hand motions, facial expressions, and body postures to communicate meaning. These non-verbal languages are exclusive to particular locations and civilizations, and no universal sign language exists that is comprehensible worldwide.

The World Health Organization (WHO) indicates that almost 466 million individuals globally, comprising 432 million adults and 34 million children, have some type of deafness or hearing impairment. This is around 6.1% of the worldwide population. Forecasts indicate that by 2050, the number of individuals impacted will exceed 900 million, approximately one in every 10 people [1]. These numbers highlight the urgent necessity for efficient communication solutions, including sign language identification and translation technologies, to facilitate interaction between hearing-impaired individuals and the wider community.

In recent years, the acknowledgment of sign languages via technology has attracted considerable attention, especially due to developments in machine learning and deep learning methodologies. Convolutional neural networks, or CNNs, demonstrated exceptional performance in image recognition applications. These deep learning models are perfect for the real-time recognition of sign language motions because they can efficiently identify characteristics and structure in visual data.

The recognition of Bengali Sign Language (BSL) earned more attention from scholars due to the large global population of Bengali speakers. Despite Bengali being the seventh most spoken language globally, with over 265 million individuals proficient in the language [2], there is a notable deficiency in research and development concerning automated BSL recognition. This gap is an opportunity to explore innovative methods that can enhance communication accessibility for the hearing-impaired Bengali-speaking population. An advanced sign language recognition technology can profoundly transform society by eliminating communication barriers. It would allow hearing-impaired individuals to interact more easily with others, fostering greater inclusion and enhancing their participation in social, economic, and political activities. This approach would facilitate their full integration into society, allowing them to contribute across diverse professions and become valuable assets to their communities.

To address the challenges in Bengali Sign Language (BSL) detection, we have developed a deep learning system adept in accurately identifying and classifying BSL characters from sign images. A major issue in this field is the limited availability of comprehensive and diverse datasets for training recognition models. To tackle this challenge, we employed data augmentation methods and artificially expand the variety of the training data set, in combination with pre-trained models. The pre-trained models, originally developed on extensive datasets for similar tasks, significantly enhance our system's performance, even when employing smaller, domain-specific datasets. We aim to improve the precision and dependability of BSL identification systems by the application of CNN models and data augmentation methods, hence enabling the development of effective solutions for the hearing-impaired community. The objective of encouraging inclusive, accessible, and technologically advanced societies is furthered by this study, which also enhances sign language recognition.

The following sections of this study are organized as follows: Section II presents a review of relevant studies on Bangla Sign Language recognition. Section III outlines the methods used for dataset collection and preprocessing techniques. Section IV defines the methodology, covering the CNN-based hybrid model utilizing transfer learning. Section V indicates the results of the experiment and performance assessment. In conclusion, Section VI wraps up the paper and examines possible directions for future work.

II. Literature Review

The study of sign language recognition is a fascinating field of research. Despite the extensive research on BdSL recognition, few studies have succeeded in rendering it practical and achievable, therefore leaving it mostly unexamined. This section delineates the progression of studies related to BdSL recognition. Md. J. Raihan et al. [3] illustrated the addition of a CNN augmented with a SE block and the use of SHAP analysis, achieving an accuracy of approximately 99.86%. Hossain et al. [4] proposed a CNN-based model for sign language recognition, achieving an accuracy of 98.75%. Islam et al. [5] created a CNN-based model and assessed its performance using 10-fold cross-validation, attaining an accuracy of 99.80%. Rafi et al. [6] proposed a VGG-19-based model for the recognition of 38 distinct classes of Bengali sign gestures, achieving a test accuracy of 89.6%. Abedin et al. [7] proposed the 'Concatenated BSL Network,' combining CNN with OpenPose for hand posture estimation, achieving 91.51% accuracy in Bangla sign language recognition. Sunanda et al. [8] applied a hybrid methodology utilizing transfer learning, achieving an accuracy of 97.89% with the implementation of VGG16 and a random forest classifier. Lipi et al. [9] developed a dataset of Bengali Sign Language words and alphabets and proposed a 10-layer Tensor Flow-based CNN for static gesture recognition, achieving 92.50% accuracy and an F1 score of 92%. Miah et al. [11] created BenSignNet, a nine-layer CNN model, which enhances generalization through the utilization of several datasets (BdSL, KU-BdSL, and Ishara-Lipi) with segmentation and augmentation approaches. The model attained remarkable accuracies of 94.00%, 99.60%, and 99.60% on the corresponding datasets.

III. Data Set Collection And Preprocessing

We utilized the KU-BdSL dataset from [11] [12], which is publicly available and comprises 1,500 samples representing 30 Bengali alphabets. The data was collected from 33 individuals, including 25 males and 8 females. Each image is 512 × 512 pixels with 8-bit RGB channels. The data set is balanced, with each class containing 50 samples. To enhance the robustness of the data set, various data augmentation techniques were applied, such as random brightness adjustments, RGB shifting, and motion blur. As described in Table I, the augmentation parameters include random brightness and motion blur. This process resulted in an augmented dataset of 7,500 samples. After feature extraction, the data set was partitioned in an 80/20 proportion for training and testing.

TABLE I
Augmentation Parameters

Augmentation Method	Parameter	Value
RGB Shift Limit	Red	5
	Green	5
	Blue	5
	Probability	70%
Motion Blur	Blur Limit	7
	Probability	70%
Random Brightness	Probability	70%

IV. Methodology

Our solution uses a convolutional neural network to extract characteristics. CNN acquires various characteristics from images during training. Furthermore, we implemented transfer learning to allocate the pre-trained weights in the model, thereby improving its overall performance. The methodological steps of our approach are illustrated in Fig. 1.

A. Convolutional Neural Networks

A Convolutional Neural Network is a bio-inspired mechanism derived from the structure of the human brain. The procedure involves a mathematical operation known as convolution, which applies to two functions: a segment of a digital image and a feature detector referred to as a filter, producing an output function termed the feature map. Various convolutional layers extract distinct feature types; first convolutional layers are recognized for capturing low-level characteristics, such as picture edges, whereas subsequent layers are renowned for capturing high-level composite features. This approach diminishes the quantity of trainable factors and effectively examines positional and orientational dependencies within an image.

ResNet-50 is a 50-layer deep convolutional neural network under the family of Residual Networks, commonly referred to as ResNet [13] . Under this construction, the network could learn residual mappings, which are much easier to optimize than the conventional raw mappings, enabling considerable gains in network depth. ResNet-50 takes an input image of size 224x224x3 (RGB) and produces feature maps or class

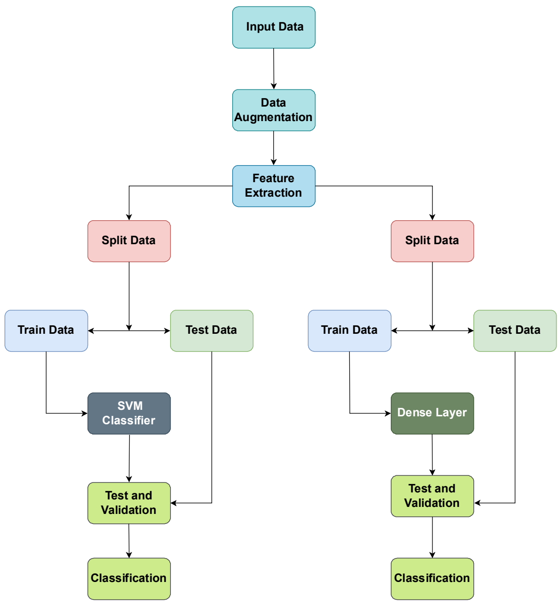

Fig. 1. Procedure of methodological steps.

probabilities, depending on the task. Having around 25.6 million parameters, ResNet-50 is extensively utilized in numerous computer vision applications. We employed ResNet-50, pre-trained on 'ImageNet' with its fully connected layers excluded, as the foundation of our Bangla sign language identification system. Figure 2 illustrates the architecture of ResNet-50.

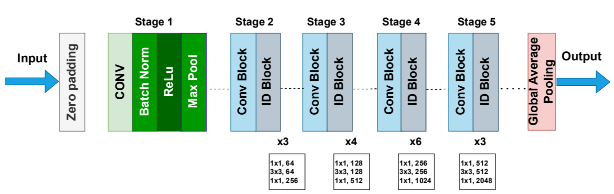

Fig. 2. ResNet-50 architecture.

B. Support Vector Machine

The Support Vector Machine (SVM) is a supervised machine learning technique extensively employed for classification and regression tasks. It is very productive for high-dimensional data and limited datasets. Support Vector Machines (SVM) seek to identify the ideal hyperplane that maximises the margin between classes, hence assuring effective generalisation to novel data.

C. Dense Layer

A customised neural network was developed with TensorFlow's Keras API for multi-class classification. The architecture consists of an input layer followed by two fully linked layers containing 256 and 128 neurones, respectively, with each employing the ReLU activation function to incorporate non-linearity. Dropout layers with a 30% rate were incorporated following each dense layer to alleviate overfitting by randomly deactivating neurones during training. The output layer employs a softmax activation function to produce class probabilities. The model was created utilising the Adam optimiser, which dynamically modifies learning rates, and sparse categorical cross-entropy for the loss function. Show in 3

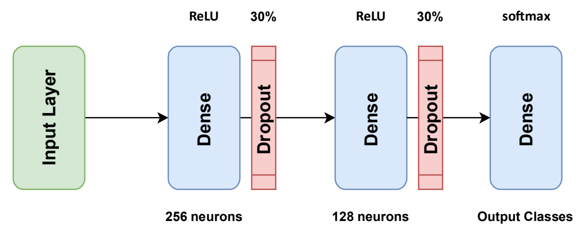

Fig. 3. Proposd Dense Neural Network architecture.

V. Results

This study applied a two-stage methodology for image classification, utilizing ResNet-50 for feature extraction and evaluating two classifiers—SVM (Support Vector Machine) and CNN (Convolutional Neural Network)—for the final classification job. The performance of two classifiers described in Table II,

TABLE II
COMPARISON OF MODEL PERFORMANCE

Methodology	Accuracy	Precision	Recall	F1-Score
ResNet-50 + SVM	**99.7%**	**0.999**	**0.999**	**1.00**
ResNet-50 + Dense Layer	99.5%	0.998	0.998	0.999

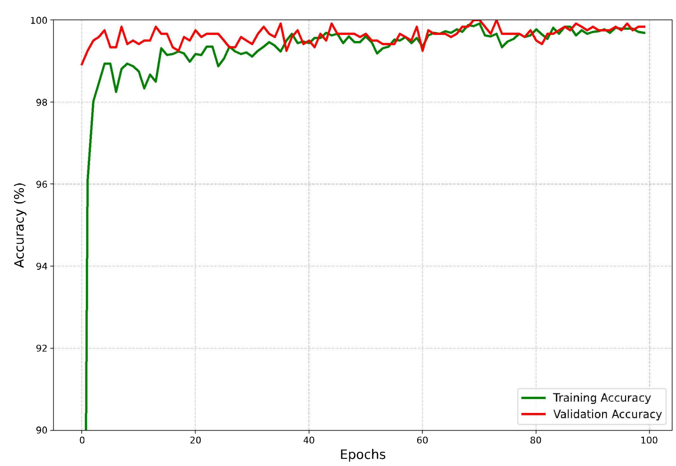

Fig. 4. Accuracy Curve (ResNet-50 + Dense Layer)

A. Result analysis

The model was trained on a dataset comprising 30 classes, with training and validation processes monitored across multiple epochs.

The graph 4 below illustrates the evolution of accuracy and validation accuracy during training:

The confusion matrix for both SVM reveals a few misclassifications across different classes. Class 0 had 1 misclassification, class 9 had 2 misclassifications, and class 29 had 1 misclassification show in 5

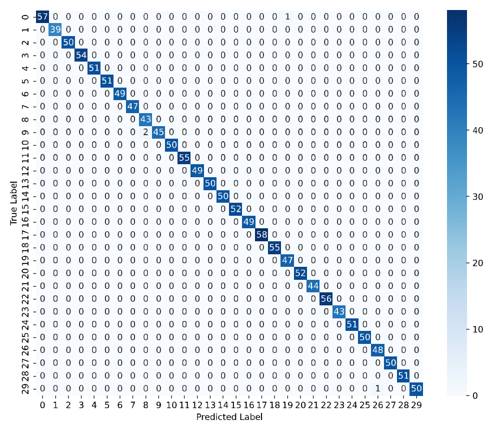

Fig. 5. Confusion Matrix (ResNet-50 + SVM)

Table III, shows a comparison with previous studies and methods.

TABLE III
COMPARISON WITH PREVIOUS STUDIES AND METHODS

Authors	Method Used	Accuracy (%)
Md. J. Raihan et al. [14]	CNN + SE Network + SHAP Analysis	99.86
Miah et al. [10]	CNN + data augmentation	99.6
Surjo et al. [15]	VGG16	98
	ResNet50	97
	MobileNetV2	95
Begum et al. [16]	Xception + Quantization	99
Proposed Model	**ResNet-50 + SVM**	99.7
Proposed Model	**ResNet-50 + Dense Layer**	99.5

B. Limitations of the study

This study uses a hybrid deep learning approach with ResNet-50 for feature extraction and two classifier to recognize Bangla Sign Language. However, limitations include a small dataset that affects generalization, lack of real-world variations, exclusion of dynamic gestures, minor misclassification of classes, and lack of a real-time user interface.

VI. CONCLUSIONS

This study successfully developed a hybrid Bangla Sign Language recognition system, demonstrating the power of combining transfer learning with machine learning classifiers. By leveraging ResNet50 for feature extraction and employing both SVM and Dense Layer classifiers, the system achieved remarkable accuracy rates on the KU-BdSL dataset. The findings highlight the system's potential to bridge communication gaps for the Bangla-speaking hearing impaired community, fostering greater social inclusion. The proposed method solves some of the big challenges in this domain due to the diversity of the data and the robustness of the model. There is still space for further improvements such as dataset expansion, enabling recognition of dynamic gestures, and deploying real-time systems on low-resource devices that could be made even more accessible and usable for a broadened set of people-a holistic, multilingual sign language recognition platform.

REFERENCES

[1] W. H. O., "Deafness and hearing loss," 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.

[2] N. Berlitz, "The most spoken languages in the world," 2021. [Online]. Available: https://scholar.google.com/scholar_lookup?title=The%20most%20spoken%20languages%20in%20the%20world&author=N.%20Berlitz&publication_year=2021.

[3] M. J. Raihan, M. I. Labib, A. A. J. Jim, J. J. Tiang, U. Biswas, and A.-A. Nahid, "Bengali-Sign: A machine learning-based Bengali sign language interpretation for the deaf and non-verbal community," *Sensors*, vol. 24, no. 16, p. 5351, 2024.

[4] S. Hossain, D. Sarma, T. Mittra, M. N. Alam, I. Saha, and F. T. Johora, "Bengali hand sign gesture recognition using convolutional neural networks," in *Proc. 2020 Second Int. Conf. Inventive Research in Computing Applications (ICIRCA)*, IEEE, 2020, pp. 636–641.

[5] M. S. Islam, M. M. Rahman, M. H. Rahman, M. Arifuzzaman, R. Sassi, and M. Aktaruzzaman, "Recognition of Bangla sign language using convolutional neural networks," in *Proc. 2019 Int. Conf. Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, IEEE, 2019, pp. 1–6.

[6] 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 the deaf and dumb community," in *Proc. 2019 IEEE Global Humanitarian Technology Conference (GHTC)*, IEEE, 2019, pp. 1–7.

[7] T. Abedin, K. S. Prottoy, A. Moshruba, and S. B. Hakim, "Bangla sign language recognition using concatenated BdSL network," 2021.

[8] S. Das, M. S. Imtiaz, N. H. Neom, N. Siddique, and H. Wang, "A hybrid approach for Bangla sign language recognition using a deep transfer learning model with a random forest classifier," *Expert Systems with Applications*, vol. 213, p. 118914, 2023.

[9] K. A. Lipi, et al., "Static-gesture word recognition in Bangla sign language using convolutional neural networks," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 20, no. 1, pp. 123–132, 2022.

[10] A. S. M. Miah, J. Shin, M. A. H. Hasan, and M. A. Rahim, "BenSignNet: Bengali sign language alphabet recognition using concatenated segmentation and convolutional neural network," *Applied Sciences*, vol. 12, p. 7452, 2022.

[11] A. A. J. Jim, I. Rafi, M. Z. Akon, U. Biswas, and A. A. Nahid, "KU-BdSL: Khulna University Bengali Sign Language dataset," Mendeley Data, 2021. [Online]. Available: https://data.mendeley.com/datasets/scpvm2nbkm/4.

[12] 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, p. 109797, 2023.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[14] M. J. Raihan, M. I. Labib, A. A. J. Jim, J. J. Tiang, U. Biswas, and A.-A. Nahid, "Bengali-Sign: A machine learning-based Bengali sign language interpretation for deaf and non-verbal people," *Sensors*, vol. 24, no. 16, p. 5351, 2024.

[15] G. S. Surjo, B. K. Ghosh, M. J. Alam, M. Razib, and S. A. Bilgaiyan, "A comparative analysis between single and dual-handed Bangladeshi sign language detection using CNN-based approach," in *Proc. 2023 Int. Conf. Computer Communication and Informatics (ICCCI)*, Coimbatore, India, Jan. 2023.

[16] N. Begum, S. S. Khan, R. Rahman, A. Haque, N. Khatun, N. Jahan, and T. QMXBdSL49, "An efficient recognition approach for Bengali sign language with quantized modified Xception," 2023.