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

Bangla Sign Language Digitization

Razi Amin Rizvee
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
raziamin98@gmail.com

Md Shahoriar Hossain Shojib
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
shojib.shojib10@gmail.com

MD Naim Akhand
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
naimdit@gmail.com

MD. Humaun Huda
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
223311043.cse.vu@gmail.com

Md Mizanur Rahman
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
itsmizan03@gmail.com

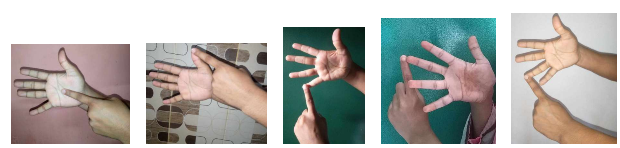
Md. Taufiq Khan
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
khantaufiq2001@gmail.com

***Abstract*—Sign language numbers, or letters made using hand gestures, are a method of communication for speech-impaired people and can be useful in many fields of technology. Current technology has advanced enough to allow 3D scanning of real objects. These required machines and software are still costly for most people and are not readily available to the public. 2D image scanning is still a great way to improve awareness and teach those who wish to learn sign language. The research aims to increase the readily available information count and contribute to creating a method, a system, an application, and a way to spread the knowledge to all. This paper has used publicly available image resources for classifying Bengali sign language letters অ(a), আ(A), ই(i), ঈ(I), and এ(e). A set of techniques was used to create a system that includes image processing techniques SMOTE (Synthetic Minority Oversampling Technique), CLAHE (Contrast Limited Adaptive Histogram Equalization), and Laplacian Filter. Deep learning CNN (Convolutional Neural Network) feature extractors from the DenseNet (Densely Connected Networks) family and the ResNet (Residual Networks) family were used in machine learning classifiers such as a custom TensorFlow Keras classifier, XGBoost (Extreme Gradient Boosting), SVM (Support Vector Machine), and Random Forest to find the best result from a limited dataset, technology, and hardware. The study aims to enhance the recognition accuracy and efficiency of Bengali sign language classification for improved accessibility and communication for the deaf and speech-impaired community.**

***Index Terms*—Bangla sign language, Convolutional Neural Network, Image processing, Image detection**

I. Introduction

Communication between people requires the involved people to understand each other's language. There is a vast gap between those who speak with their voice and those who communicate through gestures. Sign languages differ from country to country in terms of gestures, body language, and face expressions. The grammar and structure of a sentence also vary a lot. English or American sign language is one of the more universally used languages but sign language includes Bangla, Hindi, Korean, Chinese, Japanese, etc. Translating English into Bangla or Bangla into English is easy nowadays. That is mostly for spoken languages, Sign language translation is not still that popular or seen as important. The research aims to fix this issue through image processing and deep learning. More than 70 million people use sign language as means of communication.This number only increases more when it involves communication between deaf, speech impaired and those who speak with their voice. Deaf people in Bangladesh often do not have access to treatment or education, and commonly face discrimination. But technology as made it so that a single smart communication device is able connect anyone to the whole world. It is the goal of this study to make BSL more available to the deaf people of Bangladesh.



(a) অ (b) আ (c) ই (d) ঈ (e) এ

Figure 1: the Five Letters

II. Related work

Barathi Subramanian and team proposed the MOPGRU model, which captured the full information dependency in time series data with a high prediction accuracy of an average of 95 and a faster convergence speed. [1]

KASAPBAŞI and team worked with CNN models and got a result of 99.38 percent accuracy with excellent prediction ana smallll loss (0.0250) [2].

Kanchon Kanti Podder and team found that ResNet18 performed best with 99.99 percent accuracy, precision, F1 score, sensitivity, and 100 percent specificity. [3]

Shagun Katoch and team found SURF (Speeded Up Robust Features) features with (SVM) and(CNN) are used for which gave a result of an accuracy of 99.14 percent on the test data [4]

Deep Kothadiya and team used four different sequential combinations of LSTM and GRU for their own dataset IISL2020. The proposed model, consisting of a single layer of LSTM followed by GRU, achieves around 97 percent accuracy over 11 different signs. [5]

III. Methodology

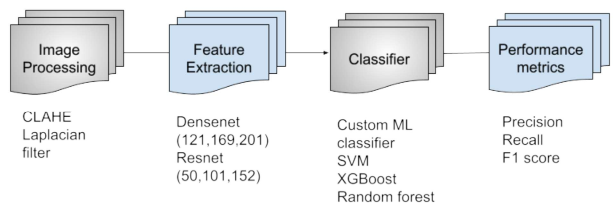


Figure 2: Methodology

This study focuses on the recognition of 5 specific letters of Bengali Sign Language (BSL). The methodology consists of the following steps: data preprocessing, feature extraction, classification, performance metrics

A. Dataset collection

letters	Classes	Image Count
অ	0	299
আ	1	319
ই	2	296
ঈ	3	300
এ	4	300
Total		1,514

Table I: Overview of the dataset for five Bengali Sign Language letters.

The dataset contains images of five Bengali vowels, organized into five classes (0 to 4). Each image is labeled according to the respective class.

B. Data processing

The dataset pre-processing included CLAHE to improve image contrast in low light and enhance local details in the images. Laplacian Filter was used to reduce noise and to sharpen the images, making the features more noticeable. The image data were resized to 224×224 pixels for compatibility with deep learning pre-trained models and normalized to the range [0, 1] after pre-processing. SMOTE is for the class imbalance due to fewer count of iamge samples between the classes.

C. Features Extractor

Deep learning CNN Pre-trained models Densenet family(121,169, 201) and Resnet variants(50,101,152) are used in a mixed combination to extract low-level and abstract features from the images for numerical value annotation and vector conversion which was transferred to the classifier.

D. Classifiers

In my system, I have applied a custom neural network classifier with TensorFlow for Bengali Sign Language letter classification. The model is built based on the feature extractor (e.g., DenseNet121). The system design includes a fully connected layer upto 512 and ReLU (Rectified Linear Unit) activation function to capture complex patterns in the extracted features. A Batch Normalization layer immediately after the fully connected layer to stabilize training and accelerate convergence. This normalizes the activations and prevents gradient explosions or vanishing during training. Overfitting was mitigated using a dropout layer with a dropout rate of 25 percentage which randomly deactivates a fraction of neurons during training. The final layer of the network is a connected output layer with five units, respective to the number of classes in the dataset. Softmax activation function was used to output a probability distribution over the five classes. The model is compiled using Sparse Categorical Cross-Entropy as the loss function and the Adam optimizer with a learning rate of 0.0001. In addition to the custom CNN other classifiers such as Support Vector Machine (SVM), Random Forest and XGBoost was tested to see if a better outcome was possible.

IV. Results

Classes	Precision	Recall	F1-Score	Support
0	0.96	0.94	0.95	52
1	1.00	1.00	1.00	51
2	0.84	0.90	0.87	52
3	0.83	0.92	0.87	52
4	0.95	0.79	0.87	52
Accuracy	**0.91**			**259**

Table II: Classification Metrics for Bengali Sign Language Letters

This study has explored the performance of various CNN pre-trained models for the classification of Bengali Sign Language. Mainly DenseNet (121, 169, 201) and ResNet (50, 101, 152) were used as feature extractors, and the extracted features were classified using a custom classifier, SVM, random forest, and XGBoost. Among these combinations, the DenseNet121 architecture paired with a custom classifier yielded the best results. The overall system shows an accuracy of 91 percent, with a range of results across different classes. Class 1 achieved perfect precision, recall, and F1-score (1.00), while other classes showed slightly lower scores. Class 0 had a strong F1-score of 0.95, while Class 4 had a lower recall of 0.79 but still maintained a solid F1-score of 0.87. Overall, the model performs well, with good balance in precision, recall, and F1 scores across most classes, although some classes have room for improvement, especially in recall.

V. Limitation and Future Work

The limitation of the study is the use of limited resources. One of the main focuses of the research was to use limited resources to get good enough results that can be used in the development of an application that can teach people. But to get better results it is necessary to use more advanced software and hardware along with a greater number of image samples. Future work must use better combinations of different techniques to get more abstract details. The study was done as an initial test to understand the result from the few specific combinations of the used techniques. The result of the study will be used as a reference for the future combination that will done on better resources. The limitations also include how to make people aware of the application and to spread it among the people where the majority of people can speak using Signs. Future work includes increasing the number of letters from the entire Bangla sign language alphabet with its proper grammar, as well as Bengali numbers. Through those results create a framework for an app or software that will function as a Sign language translator. The result of the research is leveraged for the betterment of the resources and skills necessary to use them.

VI. Conclusion

This study focused on that even with little resources and a limited dataset, a person, a student can achieve meaningful results in Bengali Sign Language (BSL) classification. By utilizing advanced image processing techniques and pretrained deep learning models, the recognition accuracy reached a commendable level, comparable to outcomes achieved by teams with access to larger datasets and better computational tools. This highlights the potential of combining innovative techniques with available resources to address real-world challenges and make a real contribution. The study highlights that impactful solutions do not always require extensive resources but rather a planned utilization of what is available, paving the way for accessible technologies that benefit underrepresented communities.

References

[1] Subramanian, B., Olimov, B., Naik, S. M., Kim, S., Park, K.-H., Kim, J. (2022). An integrated mediapipe-optimized GRU model for Indian sign language recognition. Scientific Reports, 12(1), 11964. https://doi.org/10.1038/s41598-022-15998-7

[2] Kasapbaşi, A., Elbushra, A. E. A., Al-Hardanee, O., and Yilmaz, A. (2022). DeepASLR: A CNN based human computer interface for American Sign Language recognition for hearing-impaired individuals. Computer Methods and Programs in Biomedicine Update, 2(100048), 100048. https://doi.org/10.1016/j.cmpbup.2021.100048

[3] Podder, K. K., Chowdhury, M. E. H., Tahir, A. M., Mahbub, Z. B., Khandakar, A., Hossain, M. S., Kadir, M. A. (2022). Bangla Sign Language (BdSL) Alphabets and Numerals classification using a deep learning model. Sensors (Basel, Switzerland), 22(2), 574. https://doi.org/10.3390/s22020574

[4] Katoch, S., Singh, V., Tiwary, U. S. (2022). Indian Sign Language recognition system using SURF with SVM and CNN. Array (New York, N.Y.), 14(100141), 100141. https://doi.org/10.1016/j.array.2022.100141

[5] Kothadiya, D., Bhatt, C., Sapariya, K., Patel, K., Gil-González, A.-B., Corchado, J. M. (2022). Deepsign: Sign language detection and recognition using deep learning. Electronics, 11(11), 1780. https://doi.org/10.3390/electronics11111780