Creating Multiclass Bangladeshi Sign Word Language for Deaf and Hard of Hearing People and Recognizing using Deep CNN Techniques

Md Rasel Uddin[1]

[1] Department of Computer Science and Engineering, University of Information Technology and Sciences (UITS), Dhaka 1212, Bangladesh.

Abstract

Sign language provides the solution to the communication obstacle for the community of deaf and hard of hearing. This community often faces problems communicating with normal people and among themselves, and sometimes they need an interpreter. But this is problematic and costly. As they are also vital to society, this issue should be solved automatically by recognizing sign language. Computer vision-based sign language recognition (SLR) can mitigate this issue by automatically recognizing sign language. Much research has been done on Bengali Sign language Recognition. Most of them are limited to recognizing Bengali characters and numbers. However, no significant research has been done on Bangladeshi Sign Word Language. This might have happened because of the scarcity of the Bangladeshi Sign Words dataset. In this paper, we have constructed a primary Bangladeshi Sign Language (BdSL) dataset consisting of Bangladeshi sign word images and proposed a Deep Convolutional Neural Network (DCNN) method for recognizing static Bangladeshi sign words. We recognized 11 Bangladeshi daily useful sign words from static images with high accuracy. The proposed deep convolutional neural network model shows an accuracy of 99.12% for the recognition of 11 classes of images of static signs. The model was trained on a dataset size of 1105 images which is constructed from Bangladeshi Sign language users and volunteers. This research work will help to improve the interaction between the D&D community and the general people.

1. Introduction

Communication is an essential aspect of human life. But, in our society, a community exists named the D&D community. Deaf and hard-of-hearing people are those who either can't listen or can't speak. So, they use diverse signs to communicate among themselves and others, known as sign language. Like spoken language, sign language is also very important. But it is different from spoken language. As deaf people can't communicate verbally, they have to use their body parts, especially their hands to indicate their thoughts. Sign language is all about gestures. Nearly 7% of the global population exercise sign language. In Bangladesh, around 2.6 million people use sign language [1]. It is difficult for these communities to communicate with people and among themselves. They face many problems while communicating. Sometimes the misery of their communication is unbearable as most people do not understand sign language. So, they are being ignored in society. In Bangladesh, some organizations are working for the D&D community. The Centre for Disability in Development (CDD) is one of them. But these organizations are following traditional ways of training and learning sign language is not easy. Modern technology can reduce their suffering by more than 80%. Researchers may contribute to addressing and solving this problem by automatically recognizing sign language from the gesture. An automatic sign language recognition system can significantly reduce the suffering and solve the communication barrier for the D&D community.

Around the world, there has been much research on sign language over the past few decades. Many sign languages are available in many countries such as American Sign Language (ASL), British Sign Language (BSL), Japanese Sign Language (JSL), Indian Sign Language (ISL), etc. There are differences among these sign languages. They differ from place to place. An official sign language is also recognized in Bangladesh known as Bangladeshi Sign Language (BdSL). Researchers all over the world have developed various models to recognize sign languages such as fuzzy logic, neural networks, and transfer learning. Sign language has two categories, named isolated or static sign language and continuous sign language. Lionel Pigou et al. [2] used a convolutional neural network for Italian Sign language recognition for continuous sign language of gesture. Ankita Wadhawan et al. [3] applied deep learning for ISL and Kshitij Bantupalli [4] used both deep learning and computer vision for ASL (American Sign Language) recognition of static signs. The transfer learning technique is also a popular method to recognize sign language. Authors in the paper [5] used a pre-trained model named EfficientNetB4 and achieved 98 percent training and 95 percent testing accuracy. In the paper [6], several pretrained transfer learning models such as VGG16, VGG19, ResNet50, and AlexNet were used on the Bangla sign language dataset where VGG16 achieved the highest accuracy. Support Vector Machine (SVM) which is a machine learning method was used [7] to classify Bangla sign digits and achieved an accuracy of 98.65%.

As Bangladesh has lots of D&D people, a system should be developed for BdSL (Bangladeshi Sign Language) users to remove their communication barrier. Researchers proposed many methods for Bengali Sign Language recognition. Md. Sanzidul Islam et al. [8] presented a dataset of the static sign of characters and numbers, and Ragib Amin Nihal et al. [9] used transfer learning to recognize the static sign of Bengali characters and numbers. Besides M.A. Hossen et al. [10] also presented another model for the recognition of static 37 letters in Bengali alphabets. Methods like concatenated segmentation [11] and angular loss functions [12] to recognize Bengali Sign Language Alphabet. However, most of the research is limited to recognizing characters and numbers. In this paper, we have proposed a model of a deep convolutional neural network that can recognize 11 daily useful static sign words (bad, beautiful, friend, good, house, me, my, request, skin, urine, you). The proposed framework can recognize the static sign taken as input on the fitted model. The model is fitted on 1105 images of static sign words of a primary dataset. The contribution of this paper is as follows:

- We created an original Bangladeshi Sign Language words (BdSLW) dataset constructed on images collected from different users which is the first dataset on words in Bangladesh.
- We modified the original CNN architecture for recognizing the 11 daily useful static Bangladeshi Sign words and achieved a high accuracy of 99.12%.

The rest of the paper is organized as follows: Section 2 represents the literature review, and the proposed methodology is shown in Section 3. The experimental setup is discussed in section 4. In section 5, results analysis and discussion are demonstrated. Section 6 represents the comparison of our research with existing work. Finally, the conclusion and future work is presented in section 7.

2. Literature Review

A good number of research papers have been published on Bengali Sign Language Recognition. Many researchers have researched the recognition of the Bengali Sign alphabet and numbers. The majority of research was done on the recognition of English characters, Bengali characters, English numbers, and Bengali numbers. There is very rare research done on Bengali or Bangladeshi static sign words. Abdul Muntakim Rafi et al. [13] proposed a model using a VGG19-based convolutional neural network model. The model recognizes the Bengali Sign Language of the alphabet. He collected a total of 12581 different hand signs of 38 alphabets. The model of his research achieved an accuracy of 89.6%. Lionel Pigou et al. [2] presented a paper to recognize sign language using a convolutional neural network. The model recognized 20 Italian Sign Language gestures with high accuracy. The model used Microsoft Kinect, a convolutional neural network, and GPU acceleration. The system developed in [14] by Mohammad Aminur Rahman detects 6 Bengali vowels and 30 Bengali consonants using real-time computer vision. They applied K- nearest neighbors to classify binary images of 36 Bengali Sign Characters. They used 3600 images to train the model and achieved an accuracy of 98.17% for vowels and 94.75% accuracy for consonants. Roy et al. [15] also developed a model for isolated compound characters with a benchmark. Md, Sanzidul Islam, et al. [16] used a dataset composed of 50 sign images to recognize sign language using CNN. The Bengali alphabet of sign language was recognized using a neural network ensemble by Bikas Chandra Karmokar et al. [17]. The developed model can interpret sign language into text and vice versa. They used a webcam for sign language and used an efficient neural network (NNE) to recognize sign language. Md. Sanzidul Islam et al. [8] introduced an open-access dataset of isolated characters for Bengali Sign Language, known as Ishara-Lipi. The dataset is composed of 50 sets of 36 Bengali sign Characters. Among 36 characters, 30 characters are consonants and 6 are vowels. The datasets were collected from both deaf and general volunteers. The datasets consist of 1800-character images after pre-processing. Lean Karlo S. Tolentino et al. [18] addressed a system to recognize static sign language of ASL using deep learning. Using the CNN model, they acquired 90.04% accuracy in the ASL alphabet, 93.44% for numeric, and 97.52% for static sign word recognition. Srujana Gattupalli et al. [19] also presented a model that recognizes sign language by pose estimation. A system was developed that can translate sign languages into text using SVM by Travieso [20]. Ragib Amin Nihal et al. [9] provided a system to recognize the Bangla Sign alphabet using zero-shot and transfer learning. The researchers employed both conventional transfer learning and modern zero-shot learning (ZSL) techniques for analyzing both seen and unseen data. They used a dataset of 35,149 images. Their model achieved 68.21% of harmonic mean accuracy, 91.57% of seen accuracy, and 54.34% of zero-shot accuracy. For transfer learning, an overall 93.68% accuracy was achieved. Muhammad Aminur RAHAMAN et al. [21] recognized hand-sign-spelled BdSL using a modeling algorithm. M.A Hossen et al. [10] introduced another system to recognize Bengali Sign Language using a deep convolutional neural network where they recognized 37 letters of the Bengali alphabet. The dataset was the collection of, 1147 images. The model yielded an accuracy of 96.33% on the training dataset and 84.68% on the validation dataset. A deep convolutional neural network (DCNN) served as the foundation for developing the model.

In this paper, we have recognized 11 daily useful words in Bangladeshi sign language. So far, good numbers of research have been done on the alphabet and numbers of Bangladeshi sign language. But rare research has been done on sign words, so we have created a sign word dataset and utilized a CNN model for recognizing Bangladeshi Sign Words with high accuracy.

3. Proposed Methodology

The model which we proposed in this paper is a deep learning method for recognizing Bangladeshi Sign Language (BdSL) words from static sign images with high accuracy. The model is trained on a primary image dataset. So, image dataset construction is mandatory. A primary Bangladeshi Sign Language Words dataset is constructed using a static sign image. The block diagram of the proposed method is depicted in Figure 1.

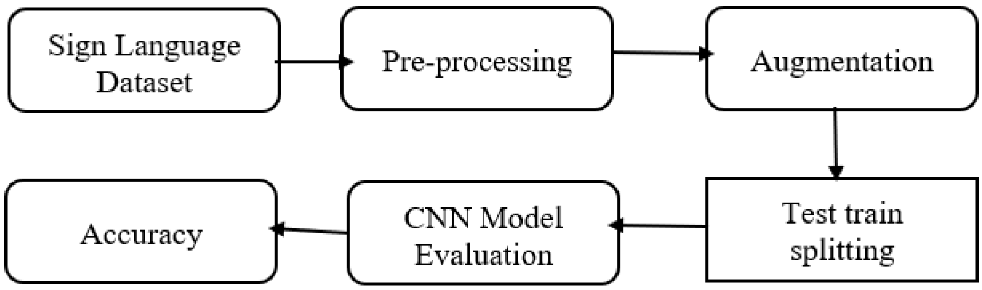


Figure 1: Block Diagram of Proposed Methodology

We have split our collected dataset into train and test sets after pre-processing and data augmentation. The splitting process was performed randomly. Then we move on to the CNN model design. Model training on train data and evaluation take place on the designed model. Finally, we give input images to the proposed model, and it provides the predicted output, recognizing the sign word.

3.1. Dataset Description

Data acquisition is a crucial part of this research as we have trained our model on a primary dataset. There is no available dataset of Bengali Sign words as per the author(s) knowledge. We have collected a total number of 1105 images of 11 classes of Bangladeshi sign words. Bangla Sign Language Dictionary is used as a reference for Bangladeshi Sign Language gestures which is published by the National Centre for Special Education under the ministry of social welfare. The data collection process is depicted in Figure 2 and the sample of our constructed dataset is presented in Figure 3.

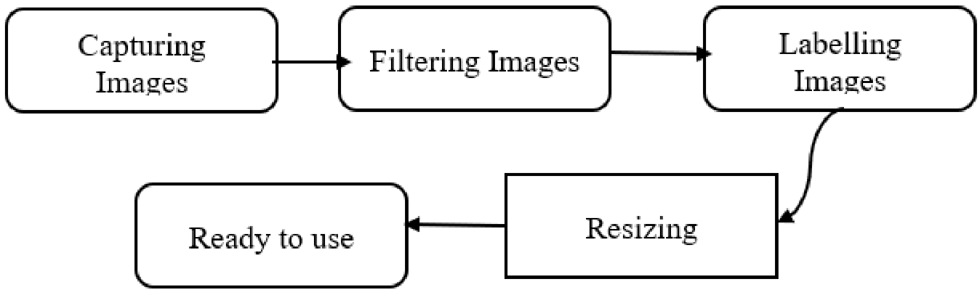


Figure 2: Data collection process.

3.1.1. Capturing Images: The images have been collected from both hearing-impaired people and general people. The pictures were captured on a Samsung Galaxy G7 camera at 60 fps and 720p resolution. Each sign was performed several times by a single signer in different lighting environments and hand postures. Besides, the background is the same for almost every image for consistency. In total, 1105 images were captured from volunteer signers. The images are captured as RGB.

3.1.2. Filtering and Labelling data: After capturing images, we labeled those images. We have classified those images as good quality and bad quality images. Bad-quality images were recaptured. We have labeled those images as the name of the word categories and separated those images into different classes or folders. Hence, there are 11 classes/folders for 11 sign words.

3.1.3. Resizing Images: As the main focus of the research is the recognition of sign words from images, images should be cropped keeping an important portion. Besides, height and width should be fit for the model. All images were resized to 224X 224. Then the dataset is ready to train.

(1) Bad (2) Beautiful (3) Friend (4) Good

(5) House (6) Request (7) Skin (8) Urine

(9) You (10) Me (11) My

Figure 3: Dataset Sample.

3.2. Pre-processing

After the development of the sign words dataset, it is vital to pre-process the dataset before training, such as rescaling and reshaping the images. The images were reshaped as 32 x 32 pixels as the model input size. The rearrangement of images is done, and we tried to keep the background of images the same to enhance the accuracy of the model.

3.3. Augmentation

The performance of a deep neural network mostly depends on datasets. Model accuracy increases with the increase in dataset size. Our dataset's initial size is 1105 images. So, we have expanded our dataset to 3835 images using the augmentation technique. Some common forms to augment data are random crops, horizontal flipping, and color augmentation [22]. Data augmentation also mostly shifts, flips, brightness, and zoom images. We have applied techniques like rotation, zooming, and flipping to augment our dataset.

3.4. Splitting the Dataset

Splitting the dataset into the training set and test set is important so that we can predict the test set of images using our trained model. We have kept 20% of images of our dataset for the test set and 80% for training our model.

3.5. CNN Model

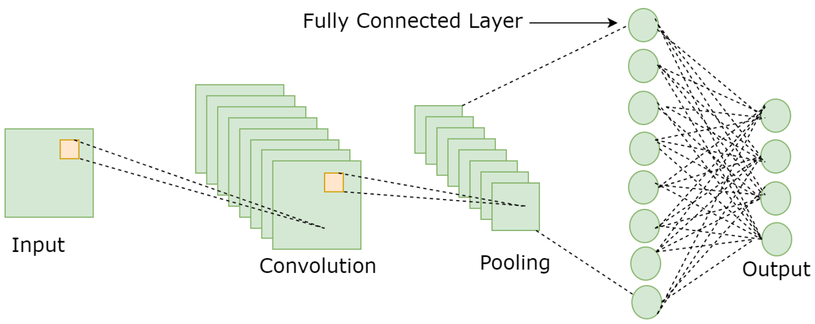


Figure 4: Basic CNN Architecture

There are several deep learning algorithms and a convolutional neural network (CNN) is one popular algorithm. This method takes an input image, allocates weights and biases to various features or objects within the image, and can distinguish between them [23]. Convolutional Neural Networks are used in an extensive limit of applications such as image and video recognition, image classification, medical image analysis, and computer vision. In CNN, the word "convolution" refers to mathematical actions of convolution that employ multiplying two functions that generate a third function which reveals how the form of one neuron is transformed by the other. There are several components of a CNN architecture. A basic CNN architecture is illustrated in Figure 4 with its basic components.

3.5.1. Convolutional layer: This is the initial layer of CNN architecture that takes input images and derives diverse features from the input. The mathematical operation between convolution and filter of a particular size M x M is performed. The output feature maps of this layer are produced through the convolution of the previous layer's feature maps with collection filters. These filters, which are the sole trainable parameters of the convolutional layer, are acquired through the back-propagation algorithm during training. [24].

3.5.2. Pooling layer: The main task of this layer is to diminish the size of the co-evolved features, reducing the computational resources required to process data, accomplished through dimensionality reduction. The main idea of pooling is down-sampling which will reduce the complexity for further layers [25]. There are two types of pooling. One name is Max Pooling and another is Average Pooling. Max pooling retrieves the highest value within the kernel-covered region of the image, while average pooling calculates the mean of all values. Additionally, max pooling functions as a noise reducer.

3.5.3. Fully Connected (FC) layer: This layer is composed of neurons as well as weights and biases. Fully Connected and CONV operations can be computed in the same way by matrix multiplication [26]. Generally, the layer connects the neurons across two distinct layers, with input images from preceding layers flattened and supplied to this layer. This stage begins the classification process. This layer takes place earlier on the output layer and forms the final layers of a CNN design. FC layer is shown in Figure 5.

3.5.4. Dropout: Dropout is mostly used to solve overfitting problems. In the context of deep neural networks, dropout's averaging effects can be described by three recursive equations, which involve an estimation of expectations using normalized weighted geometric means [27]. The dropout layer is demonstrated in Figure 5. Since a fully connected layer is linked to all features, this can potentially lead to overfitting on the training dataset. To address this problem, some neurons are intentionally excluded from the neural network during the training process. When the input is small, this also creates an overfitting problem.

3.5.5 Activation Function: The activation function is among the most crucial parameters of the CNN model. Activation functions are utilized to acquire and approximate intricate and continuous relationships between variables within the network. Numerous activation functions exist, including ReLU, Softmax, tanh, and Sigmoid, among others. The improved activation function provides faster convergence and also neural network identification accuracy improves [28].

4. Experimental Setup

The system we have presented here is based on a deep Convolutional Neural Network (CNN) that is implemented in a device of processor AMD Ryzen7, 3700X 8-Core Processor 4.20 GHz, 8.00 GB RAM, and a configuration of NVIDIA GeForce GT 710 2 GB graphics card. The model is a Convolutional Neural Network (CNN). Figure 5 illustrates the conceptual framework of the proposed CNN model architecture.

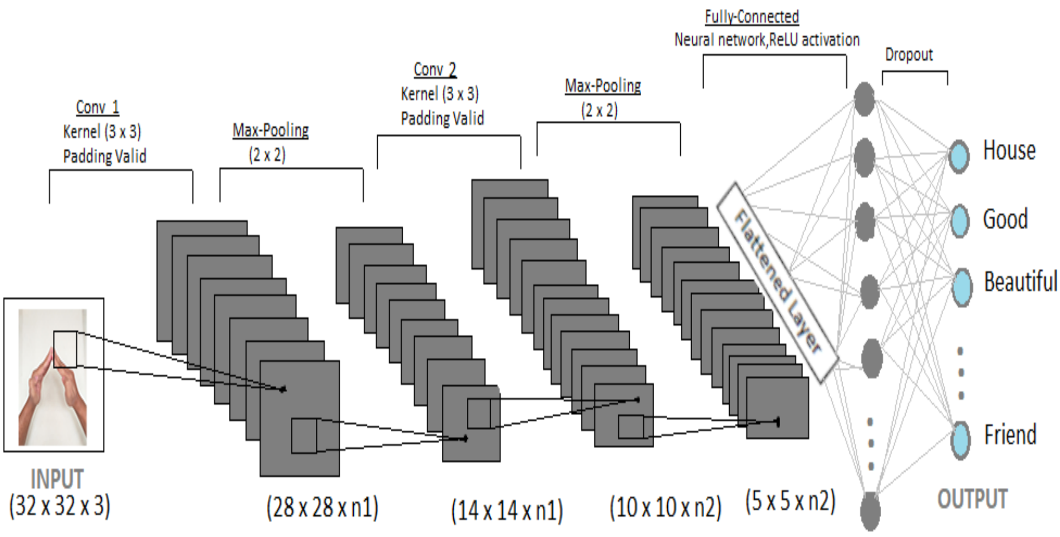


Figure 5: Proposed CNN model architecture.

The convolutional layer of our CNN model consists of 32 filters, each of which has a 3 × 3 kernel. The shape of our input images is 32 × 32 × 3. A pooling layer is taken off 2 x 2 which reduces spatial dimensions to 64 × 64. The activation function is ReLU. The dropout (0.5) function is used to overcome the overfitting problem. The model is then flattened and the fully connected layer is addressed by a dense layer along with rectified linear activation (ReLU). The model is finished with the softmax classifier to predict the probabilities of 11 Bangladeshi Sign Words. The loss function squared hinge is used an Adam optimizer is also been used in this model. Performance metrics are accuracy. Table 1 presents hyperparameters of the proposed experimental setup.

Table 1: Hyperparameters of the TL model

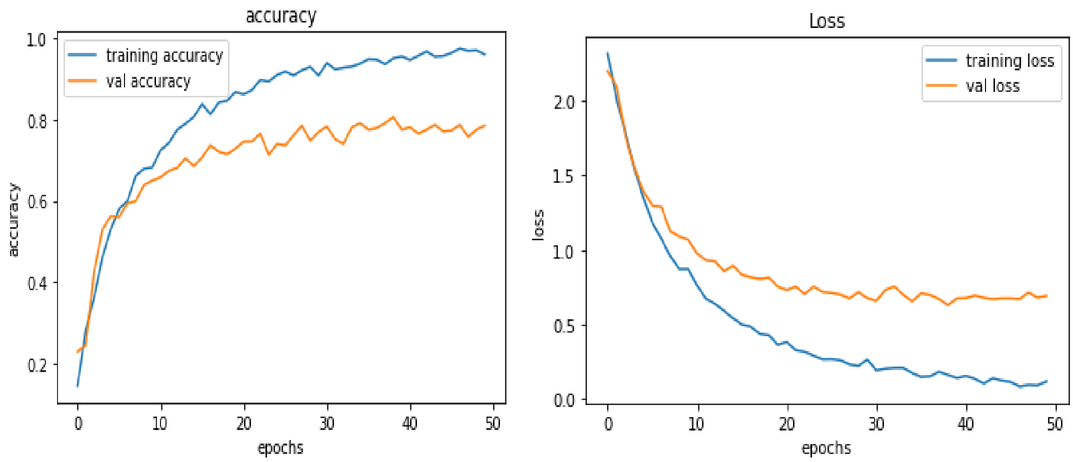
Parameters	Parameters value
Batch size	224
Optimizer	Adam
Learning rate	0.001
Epochs	50
Criterion	Cross Entropy Loss

5. Result Analysis and Discussion

In a recognition system, the accuracy rate of a model is the greatest concern. The model we have presented here is a CNN model. Our proposed model has generated 99.12% accuracy. The testing accuracy is 84.67% and the validation accuracy is 80.54%. The model is fitted for 50 epochs. The accuracy is very good compared to the size of our dataset. Table 2 represents the accuracy of our proposed trained model.

Table 2: Accuracy of the model

Parameters	Parameters value
Training	99.12%
Validation	80.54%
Testing	84.67%



(a) Accuracy **(b)** Validation Loss

Figure 6: Model Accuracy and Validation Loss

In figure 6(a), the accuracy of our proposed CNN model is demonstrated. It shows the fluctuation of both training and testing accuracy. Both training and validation accuracy increased gradually after the first epoch which finally results in an accuracy of 99.12% and 80.54% respectively after 50 epochs. Figure 6(b) shows the graph of both training loss and validation loss of our proposed model which depict that both of them decreased over epochs. The final training loss is 0.0281 and the final validation loss is 0.9458.

Grad-Cam visualization is used to validate our proposed model performance. Gradient-weighted Class Activation Mapping (Grad-CAM) utilizes gradients of a targeted concept, passing through the last convolutional layer, to generate a rudimentary localization map that accentuates significant areas in the image for concept prediction. [29]. We have used Grad-Cam to see where our model is focusing to classify the sign words of Bangladeshi sign language. Grad-cam is a visualization technique by generates a heatmap for a given class. Here we have generated images by GradCam for validation of our model. Figure 7 shows that the model is performing well, focusing on points.

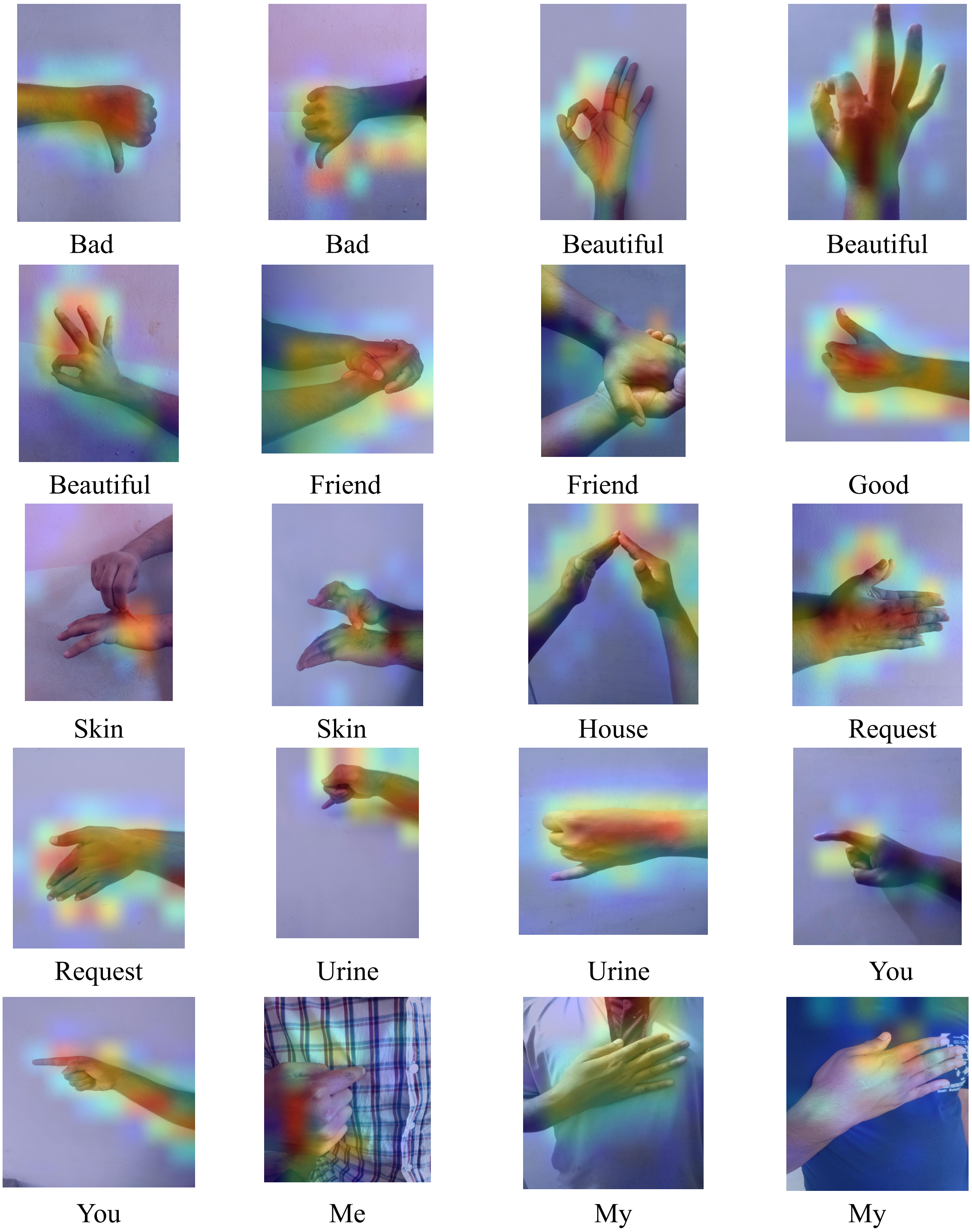


Figure 7: GradCam Visualization of the model.

The standard approach for evaluating classification models encompasses metrics such as Confusion Matrix, Precision, Recall, F1 Score, and Accuracy. We have tested our model for every 11 sign words. Table 3 presents the accuracy, correct recognition, and incorrect recognition rate of the static word. The accuracy rate was calculated using the formula (1). The formula demonstrates that the total predicted correct recognition will be divided by the multiplication of the total number of users and the total number of trials.

$$Accuracy = \frac{Total\ number\ of\ correct\ recognized\ words\ from\ users}{(Total\ number\ of\ users)(No.\ of\ Trials)} \quad (1)$$

$$Recall,\ R\ = \frac{TP}{TP + FN} \quad (2)$$

$$F1\ Score,\ F1\ = \frac{2\ x\ TP}{2\ x\ TP + FP + FN} \quad (3)$$

Equation (2) represents the formula of recall. To calculate Recall (R), divide True Positive (TP) by the sum of True Positive (TP) and False Negative (FN). Accuracy denotes the count of correctly classified samples within the entire set of samples [30]. The formula of recall and precision is presented in Equation (2) and Equation (3) where TP means true positive, FP is false positive, and FN means false negative. In equation (3), the formula for calculating the F1 score is presented. Precision (P) pertains to the number of accurately identified samples among the total number of samples identified. On the other hand, recall (R) refers to the number of correctly identified samples among all positive samples [31]. The F1-score represents the harmonic mean of the recall and precision components. Table 3 shows the evaluation metrics of the proposed model. It shows the precision, recall, and f1-score of every 11 sign words. Besides correct recognition (CR) and incorrect recognition (ICR) of trials are given with accuracy.

Table 3: Accuracy Rate of Sign Words.

Static Word	Precision	Recall	F1-score	CR	ICR	Accuracy
Bad	0.89	0.97	0.93	8	2	80%
Beautiful	0.89	0.98	0.93	7	3	70%
Friend	0.80	0.86	0.83	6	4	60%
Good	0.82	0.84	0.83	9	1	90%
House	0.97	0.94	0.95	7	3	70%
Me	0.72	0.87	0.79	5	5	50%
My	0.87	0.78	0.82	6	4	60%
Request	0.77	0.62	0.69	7	3	70%
Skin	0.92	0.86	0.89	6	4	60%
Urine	0.76	0.76	0.76	8	2	80%
You	0.91	0.78	0.84	9	1	90%

Table 3 shows that sign words ‘good’, ‘bad’, ‘you’, and ‘me’ achieved the highest accuracy with ‘good’, and ’you’ achieving the highest among them at about 98.00%, and skin getting the lowest accuracy of 70%. The confusion matrix of our model given below shows presents the performance of the model. This confusion matrix is for 11 classes of our 11 Bengali sign words. The Confusion Matrix, as the name implies, provides a matrix output that describes the model's overall performance. A confusion matrix of multi-class classification is shown in Figure 8.

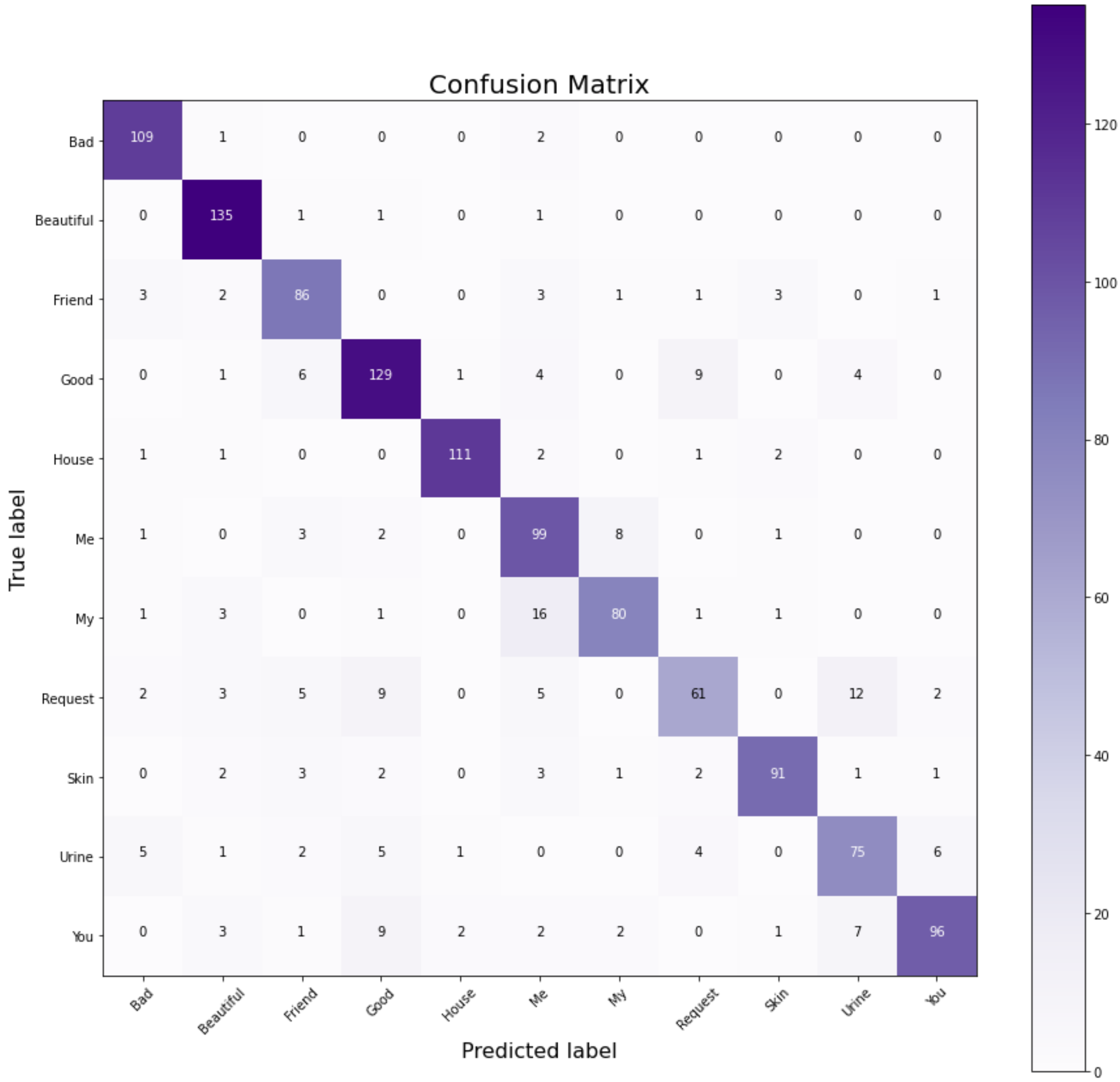


Figure 8: Confusion Matrix of the model.

The confusion matrix of our proposed CNN model is depicted in Figure 8 which describes the correct and incorrect recognition of the trained model of the training data. The model correctly recognized the ‘Beautiful’ class most and the ‘Request’ class is recognized correctly less compared to the other classes. Each class has some incorrect recognition. The ‘Request’ class demonstrates more false positive (FP) and false negative (FN) values in Figure 8. In this experiment, the result performance can be improved. The accuracy score can be higher by following some techniques such as improving image quality. If the images were taken with better pixels the camera, the accuracy would be higher. Besides, if the background of images doesn’t conflict with skin color, it will also improve accuracy and a clear image rather than a dark image can affect model performance.

6. Comparison with existing work

As sign language is a prominent problem that should be resolved, a good deal of research has been done on this issue. Table 4 summarizes several papers related to this research work.

Table 4: Summary of publications using CNN for Bangladeshi Sign Language Recognition

SN	Reference	Dataset	Method	Accuracy
01	[32]	BdSL Alphabet	Xception	98.93%
02	[33]	BdSL characters	DCNN	99.22%
03	[34]	BdSL alphabet	CNN	99.22%
04	[35]	BdSL alphabets and Numbers	CNN	99.83%
05	[8]	BdSL Characters	Dataset Construction	1800 images
06	[36]	BdSL digits	CNN model	95%
07	[37]	Bengali numbers	CNN	99.8%
08	[10]	Bengali alphabet	DCNN	96.33%
09	[1]	BdSL Characters and Numbers	CNN	98.75%
10	[13]	BdSL alphabets	VGG19	89.6%
11	Proposed	Bengali Sign Words	DCNN	99.12%, 1105 images

Table 4 shows that almost all research has worked on similar datasets using CNN or transfer learning. They either work on BdSL characters, alphabets, or numbers. But BdSL words were missed which should be introduced. So, in this paper, the issue is introduced with a good dataset construction.

7. Conclusion and future work

The main focus of this paper is recognizing sign words from static sign images. We have constructed a large image dataset of Bengali Sign Language Words and proposed a CNN model to recognize the Sign Words. The proposed CNN model has achieved 99.12%, 84.67%, and 80.54% accuracy for the training dataset, testing, and validation respectively to recognize the Bangladeshi Sign Language (BdSL) words. The model is then tested against 5 users with 10 trials, which shows the good performance of the model to recognize the 11 sign words. The major challenge of the model performance is the background of the image. The image's background color, light, and quality greatly impact model accuracy. Besides, the accuracy also proportionally depends on the size of the dataset. The research is beneficial for the D&D community of our society. It will help their communication barrier. In our future work, we will collect the dataset of continuous Bengali sign words and develop a model that can recognize the words from the video. A mobile-based and web-based application will also be developed based on the model to make the model more friendly to the users. The deaf and hard-of-hearing community can easily use the application.

Data Availability

The data used in this research can be accessed with the author's permission.

Funding Statement

This research is not funded.

References

[1] S. Hossain, D. Sarma, T. Mittra, M. N. Alam, I. Saha, and F. T. Johora, "Bengali Hand Sign Gestures Recognition using Convolutional Neural Network," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 636-641, 2020.

[2] L. Pigou, S. Dieleman, P. J. Kindermans, and B. Schrauwen, "Sign language recognition using convolutional neural networks," In European conference on computer vision, Springer, Cham, pp. 572-578, 2014.

[3] A. Wadhawan, and P. Kumar, "Deep learning-based sign language recognition system for static signs," Neural computing and applications, vol. 32, no. 12, pp. 7957-7968, 2020.

[4] K. Bantupalli, and Y. Xie, "American sign language recognition using deep learning and computer vision," In 2018 IEEE International Conference on Big Data (Big Data), pp. 4896-4899, 2018.

[5] Z. Mohammed, Y. A. Alotaibi, D. Koundal, Y. Guo, and M.M. Elahi, "Sign Language Recognition for Arabic Alphabets Using Transfer Learning Technique," Computational Intelligence and Neuroscience Article ID 4567989, 2022.

[6] M.M. Islam, M. R. Uddin, M.d Nasim AKhtar, and KM Rafiqul Alam, "Recognizing multiclass Static Sign Language for deaf and dumb people of Bangladesh based on transfer learning techniques and the primary sign word dataset," Informatics in Medicine Unlocked, 101077, 2022.

[7] S.M. Rayeed, G.W. Akram, S. T. Tuba, et al., "Bangla sign digits recognition using depth information," In Fourteenth International Conference on Machine Vision (ICMV 2021), vol. 12084, pp. 190-199, 2022.

[8] 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 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), pp. 1-4, IEEE, 2018.

[9] R. A. Nihal, S. Rahman, N. M. Broti, and S. Ahmed Deowan, "Bangla sign alphabet recognition with zero-shot and transfer learning," Pattern Recognition Letters, vol. 150, pp. 84-93, 2021.

[10] M. Hossen, A. Govindaiah, S. Sultana, and A. Bhuiyan, "Bengali sign language recognition using deep convolutional neural network," In 2018 joint 7th international conference on informatics, electronics & vision (iciev) and 2018 2nd international conference on imaging, vision & pattern recognition (icIVPR), pp. 369-373, 2018.

[11] A.S. M. Miah, J. Shin, M. A. M. Hasan, and M. A. Rahim, "BenSignNet: Bengali Sign Language Alphabet Recognition Using Concatenated Segmentation and Convolutional Neural Network," Applied Sciences 12, no. 8, 3933, 2022.

[12] S.K. Youme, T.A. Chowdhury, A. Hossain, et al., “Generalization of Bangla Sign Language Recognition Using Angular Loss Functions,” IEEE Access, vol. 9, pp. 165351-165365, 2021.

[13] 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 2019 IEEE global humanitarian technology conference (GHTC), pp. 1-7, 2019.

[14] M. A. Rahaman, M. Jasim, M. H. Ali, and M. Hasanuzzaman, “Real-time computer vision-based Bengali sign language recognition,” In 2014 17th international conference on computer and information technology (ICCIT), pp. 192-197, 2014.

[15] S. Roy, N. Das, M. Kundu, and M. Nasipuri, “Handwritten isolated bangla compound character recognition: A new benchmark using a novel deep learning approach,” Pattern Recognition Letters, vol. 90, no. 15, pp.15-21, 2017.

[16] M. S. Islam, S. S. S. Mousumi, A. S. A. Rabby, and S. A. Hossain, “A simple and mighty arrowhead detection technique of Bangla sign language characters with CNN,” In International Conference on Recent Trends in Image Processing and Pattern Recognition, pp. 429-437, 2018.

[17] B. ChandraKarmokar, K. Md. Rokibul Alam, and M. Kibria Siddiquee, “Bangladeshi sign language recognition employing neural network ensemble,” International journal of computer applications, vol. 58, no. 16, pp. 43-46, 2012.

[18] L. K. S. Tolentino, R. O. S. Juan, M. A. B. Pamahoy, Joni, R. R. Forteza, and X. Garc´ıa, “Static Sign Language Recognition Using Deep Learning,” International Journal of Machine Learning and Computing, vol. 9, pp. 821-827, 2019.

[19] S. Gattupalli, A. Ghaderi, and V. Athitsos, “Evaluation of deep learning based pose estimation for sign language recognition,” In Proceedings of the 9th ACM international conference on pervasive technologies related to assistive environments, pp. 1-7, 2016.

[20] M. C. Travieso, B. A. Jesús, and A.F. Miguel, “Sign Language to text by SVM,” In Seventh International Symposium on Signal Processing and Its Applications, Proceedings., vol. 2, pp. 435-438, 2003.

[21] 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 of Computer Science, vol. 14, no. 3, pp. 1-20, 2020.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2017.

[23] S. Ankita, and P Singh, “Image Classification: A Survey,” Journal of Informatics Electrical and Electronics Engineering (JIEEE), vol. 1, no. 2, pp. 1-9, 2020.

[24] B. Klein, L. Wolf, and Y. Afek, “A dynamic convolutional layer for short range weather prediction,” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4840-4848, 2015.

[25] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” In 2017 international conference on engineering and technology (ICET), pp. 1-6, 2017.

[26] W. Ma, and J. Lu, “An equivalence of fully connected layer and convolutional layer,” arXiv preprint arXiv:1712.01252, 2017.

[27] P. Baldi, and P. J. Sadowski, “Understanding dropout,” Advances in neural information processing systems, vol. 26, 2013.

[28] G. Lin, and W. Shen, “Research on convolutional neural network based on improved relu piecewise activation function,” Procedia Computer Science, vol. 131, pp. 977–984, 2018.

[29] R. R. Selvaraju, M. Cogswell, A. Das et al., “Grad-cam: Visual explanations from deep networks via gradient-based localization,” In Proceedings of the IEEE international conference on computer vision, pp. 618-626, 2017.

[30] M.M. Islam, M.A. Kashem, and J. Uddin, “Fish survival prediction in an aquatic environment using random forest model,” IAES International Journal of Artificial Intelligence (IJ-AI), vol.10, no. 3, pp. 614-622, 2021.

[31] M. B. Hossain, S. H. S. Iqbal, M. M. Islam, M. N. Akhtar, and I. H. Sarker, “Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images,” Informatics in Medicine Unlocked, vol. 30, pp. 100916, 2022.

[32] 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 2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), pp. 1-5, 2019.

[33] D. Tasmere, and B. Ahmed, “Hand gesture recognition for bangla sign language using deep convolution neural network,” In 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), pp. 1-5, 2020.

[34] 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 2020 IEEE Region 10 Symposium (TENSYMP), pp. 1303-1306, 2020.

[35] M. S. Islalm, M. M. Rahman, M. H. Rahman et al., “Recognition bangla sign language using convolutional neural network,” In 2019 international conference on innovation and intelligence for informatics, computing, and technologies (3ICT), pp. 1-6, 2019.

[36] 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,” Procedia computer science, vol. 143, pp. 611-618, 2018.

[37] F. M. M. Shamrat, S. Chakraborty, M. M. Billah et al., “Bangla numerical sign language recognition using convolutional neural networks,” Indonesian Journal of Electrical Engineering and Computer Science 23, no. 1 pp. 405-413, 2021.

[38] El-Komy, Amir, Osama R. Shahin, Rasha M. Abd El-Aziz, and Ahmed I. Taloba. "Integration of computer vision and natural language processing in multimedia robotics application." Information Sciences Letters, 7 (2022): 6.

[39] Khalaf, Mohamed HR, Zeinab M. Abdel Azim, W. H. A. H. Elkhateeb, Osama R. Shahin, and Ahmed I. Taloba. "Explore the E-learning management system lower usage during COVID-19 pandemic." Information Sciences Letters 11, no. 2 (2022): 537-548.

[40] Taloba, A.I., 2022. An Artificial Neural Network Mechanism for Optimizing the Water Treatment Process and Desalination Process. Alexandria Engineering Journal, 61(12), pp.9287-9295.

[41] Rayan, Alanazi, Ahmed I. Taloba, Abd El-Aziz, M. Rasha, and Amr Abozeid. "IoT enabled secured fog based cloud server management using task prioritization strategies." International Journal of Advanced Research in Engineering and Technology 11, no. 9 (2020).