Proceedings of the Second International Conference on Inventive Research in Computing Applications (ICIRCA-2020)
IEEE Xplore Part Number: CFP20N67-ART; ISBN: 978-1-7281-5374-2

Bengali Hand Sign Gestures Recognition using Convolutional Neural Network

Sohrab Hossain
Department of Computer Science and Engineering
East Delta University
Chittagong , Bangladesh
sohrab.h@eastdelta.edu.bd

Dhiman Sarma
Department of Computer Science and Engineering
Rangamati Science and Technology University
dhiman001@yahoo.com

Tanni Mittra
Department of Computer Science and Engineering
East West University,
Dhaka, Bangladesh
mittra.cse.buet12@gmail.com

Mohammad Nazmul Alam
Department of Computer Science and Engineering
Royal University of Dhaka
Dhaka, Bangladesh
mna235@yahoo.com

Ishita Saha
Department of Computer Science and Engineering
University of Science and Technology Chittagong
ishita_saha05@yahoo.com

Fatema Tuj Johora
Department of Computer Science and Engineering
University of Science and Technology Chittagong
inna.johora@gmail.com

***Abstract*- Sign language and spoken language, both of them are generated from the same human brain but their linguistic and physical transmission varies greatly from each other. The deaf people have a hearing disability and the dumb or mute people have a speaking disability except other things are normal for them. Persons having hearing and speaking impairment are often incapable of communicating their statements appropriately. So they use sign language to communicate with each other and with the rest of the world. As a result, sign language recognition (SLR) has become one of the most interesting topics in computer vision and machine learning recently. Researchers are trying to improve this language to use in a large-scale though it is not an international language. This paper proposed a model to recognize Bangla sign language gestures using convolutional neural networks (CNN) to support Bengali natives. A large publicly available sign language dataset was used for detecting Bangla Sign Language (BSL). CNN was used to recognize and classify hand image in the screen and then to categorize the hand skeletal features extracted from the image into a standard communicative meaning for all. The proposed method shows 98.75% accuracy.**

Keywords—Hand Sign Gestures; Sign Language Detection; Artificial Intelligence; Machine Learning; Deep Learning; Convolutional Neural Network

I. INTRODUCTION

Sign language is a natural linguistic for all though it is not an international language. It works as an interpreter to make a bond among the hearing weakened and ordinary people. People use signs through making body-signs, moving hands, or facial languages as a relaxed medium to convey a message. People, particularly those who are both unable to hear and speak, need to learn sign language for communication. But most of the common people do not understand sign languages [1].

The harsh truth is that deaf communal inclines to be inward-looking and fairly separate from the rest of civilization. Signers frequently have to select pantomimic gestures or written words to connect their needs. An automatic sign language conversion system will help to breakdown this statement barrier. Preferably such a structure should allow signers to use their natural sign language. A system was developed a system to analyze video clips where various gestures are used as input, and audible instructions heard as output. Angles of different parts of hand with the body were measured by examining captured pictures from input videos. The input was stored in a database with for audio sense. Convolutional neural network is used to develop feature removal models that were confirmed to be effective for picture recognition. Now the methods are being used by Google, Facebook, Amazon, etc [2-4].

Convolutional Neural Network (CNN) stirred by the visual cortex of the humanoid brain. The artificial neurons in a CNN will attach to a limited area of the pictorial field, named a receptive field. Sign language varies from country to country. American Sign Language is considered as International Sign Language among. In Bangladesh, an official sign language has been recognized freshly [5].

The Centre for Disability in Development (CDD) had taken the ingenuity to regulate connections with sign languages in 2000. There were dissimilar local variations in BSL and no national language was taught in the Training Centre. Now, BSL is constructed from handed of the core standing sign alphabets and number signs. Many researchers in BSL have similar recognition patterns of various sign languages. This paper specifically focuses on BSL detection. It will identify an easy communication scope between ordinary people and hearing-impaired people to understand the Bengali Language, and to recognize Bengali alphabets and numerals

which will beneficial for the hearing or speaking impaired people with others [6-10].

II. Related Research

For categorizing selfie sign language signs machine learning algorithms [11, 29-31] like Convolutional Neural Network architecture is widely used. The CNN is consists of three convolutional layers. To advance the speed and accuracy in detection, different sifting gap sizes is measured for individual layer. After applying a stochastic merging method that pools the recompenses of max and mean methods, a selfie sign language database is created which consists of 200 ISL gestures symbol in 5 users reliant on viewing angles for 2 seconds individual at 30fps creating a sum of 30,0000 sign gesture videos. To grasp the strength of vast training modes that is essential for convolutional neural network the training is completed in various batces. Training and authentication loss the proposed convolutional neural network architecture is experimental [12-16].

In other research, a 2D video has been taken using a cellular phone selfie camera to identify Indian sign gestures.The target of the research was to simulate algorithms which will preferably execute on a mobile platform, although the development of a mobile app was far from certainty. The predominant module became to excerpt facts frames to reduce the involvement of video facts according to frame. The prototypical works properly for non-stop video backgrounds and we are capable of perimeter our paintings to this feature video set [17].

Researchers developed a Bangla sign language system that used SIFT feature extraction and convolution neural network (CNN) for the organization. SIFT features rise the accuracy of the convolutional neural network for spotting Bangla Sign Language. This paper included 38 Bangla signs, which can be used in 51 Bangla letters. This sign language was gathered from Bangla Sign Language Dictionary and did not include a two-handed sign as this is the primary time CNN was useful to Bangla sign language. The number of exercise and testing samples was 1700 in total while it seems very high, for good accuracy and for real-life testing much greater dataset is required. But this is a small step in the direction of applying the Convolutional Neural Network. We can see when SIFT is allowed, there is a vast quantity of accuracy upsurge.This is due to the fact that SIFT helps Convolutional Neural Network to cut features without scrambling,rotation, resolution problem. Other researchers suggest a technique for sixteen predefined sign recognition with HOG structures. Since big dataset & computational efficacy, SVM's a well-organized method for result creation. HOG techniques sideways via audio-output lets in the language lessened to connect more professionally,even though many researches has been done for hand synmols recognition and detection. Fundamental idea is to use HOG feature based hand sensor for Bangla sign language recognition. The conclusion trying to expose that our process exhibits a very good presentation in testing. Though, this approach is appropriate for static sign recognition. For larger accuracy of recognition the quantity of trained photos can be increased [18-20, 27].

By right stroking and alignment of hand on proper lights the gratitude percentage of symptoms can also be improved .Again, from the source, projected a system to identify Bangla Sign Symbols. the target goal is proficiently happened [21-24, 28].In Bangla sign language recognition structures ,Freeman Chain Code for feature abstraction and ANN for credentials is a diverse method that is very rarely used. For this deliberate structure, approximately 20 symbols have been verified and maximum of them are documented positively. The planned system is conquered 96.5% reputation percentage. On or after one more investigation, computer visio based Bangla sign words appreciation structure is planned in wherever "contour analysis" is offered [25]. To find the region of interest (ROI), "Haarlike" feature-based cascaded classifier is used. "Canny edge detector" has been used after preprocessing to citation contours. Finally Sign words are detected based on the extreme computation over trials and predefined exercise contour prototypes using Inter Correlation Function(ICF) [26].

III. Methodology

In this research, a computer vision-based learning technique was proposed to notice hand signs in various decorations. To detect various sign gestures, we apply machine learning method which contains training classifier with HOG structures. We have selected CNN algorithm for forecasting targeted signs.

A. Data cleaning and data pre-processing

In this process, data were cleaned, rearranged, and modified into readable for the algorithms. Pre-processing reshaped and rescaled raw data to fit into the neural network. Particularly cleaning reduces noise in the raw data.

B. Feature extraction

Unnecessary data were normalized from the raw images to extract the main features. Feature Extraction aims to reduce the number of features in a dataset and can be implemented in Python using Keras application programming interface (API). An API is used to integrate link among a third-party platforms for access its features and services. Keras is an neural networks based API and capable of running on top of TensorFlow. We can use Keras if you need a deep learning library which is easy and fast prototyping. Here we are extracting the features we need. we took those features

which represent our images. we find which part is important from the image data. Then we find the correlation between the data. we extracting the features here.

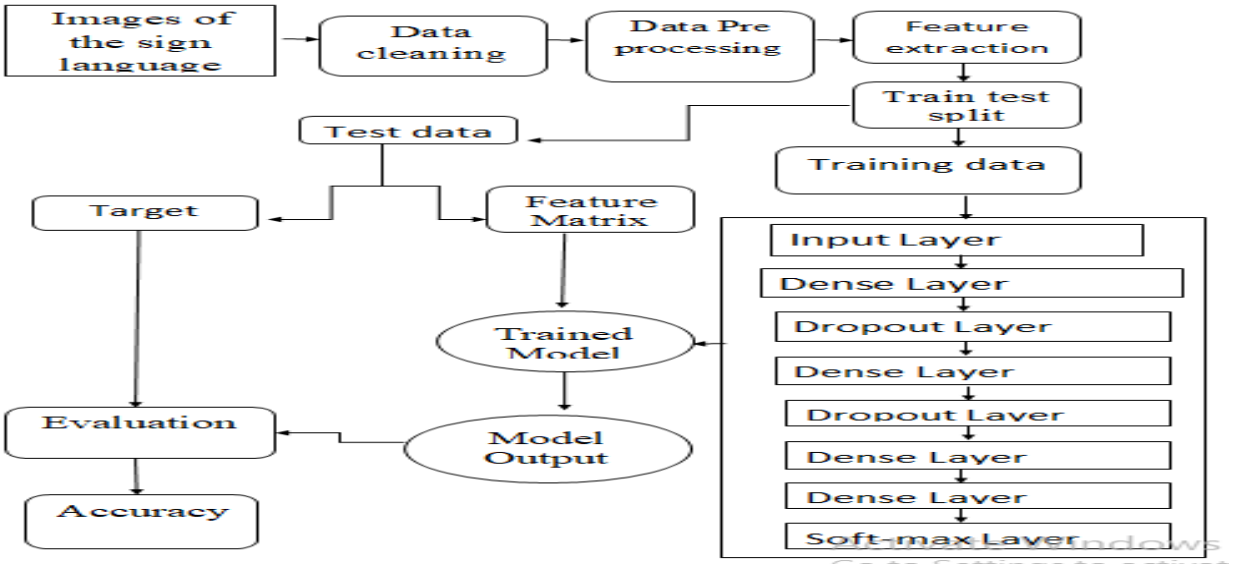

Fig. 1 The architecture of the proposed system

As an example, suppose we take a picture of a human being and there is a wall behind him.at this point, feature extraction works extracting the main & important part like the human being's face, body, etc. but not the wall cause this is not that important. So feature extraction works like this. If we don't do the feature extraction it will calculate the data which we left behind and it will cost our accuracy.

C. Train test split

In this step, we tested the model that we trained. We took 25000 images and predicted the result. The images were split into 80% for the train data and 20% for the test data.. The reason for splitting the dataset into 80% and 20% is to train the model in a neural network. The training and test data contained images and the targets. We converted the images into pixels called a feature matrix. We separated the target from the feature matrix. The neural network can calculate a huge amount of features. Data for train purposes passed through eight layers in the training phase.

D. Convolutional Neural Network (CNN).

Deep learning models are abstrusely stimulated by data dealing out and outlines in genetic nervous structure, yet have various variances from the physical and functional properties of the genomic brain which make them mismatched with neuroscience signs. It is a common term for multi-layer NN. CNN is a type of Feed-Forward - Artificial Neural Networks that is known to be vastly powerful within the field of picture sorting & credentials

1) Different layers of CNN are:

(i) Convolution: - The utmost use of the Convolution process in the situation of a CNN is to identify appropriate features on or after the image which is an involvement for the main layer. Three-dimensional inter-relation of pixels is preserved by this. It is completed by using small squares of image by attainment of picture features. (ii)Non-linearity: - ReLu is a Nonlinear procedure known as Rectified Linear Unit. ReLU is a procedure that is completed per pixel which then exceeds all the negative figures of every pixel by zero in a feature map.

(iii) Max-Pooling: - It is a discretization procedure which is based on samples. The aim is to down-sample an input illustration. It can reduce dimensionality and allow for expectations to be made about features limited in the sub-regions binned. It is additionally termed down test group helps in reducing the scale of each feature map. As soon as pooling is done, eventually, the 3D feature map will be reestablished to the Minimum 1D featured vector.

(iv) Fully Connected layer (Classification): - In these layers. All of the inputs come from One layer also connected to Every other activation unit of the Following layer. In the maximum current machine learning model, last some layers are fully connected layers which collect the data mined by previous layers, so that it can make a final output. This is 2nd most Time-consuming layer after the Convolution.

E. Images of the sign language

At first, we collected our training dataset to make a good dataset. So, all the raw images here which were taken by us for this particular research were collected from some deaf and mute persons, to know how they communicate with persons like them. We collect the sign as an image.

Fig. 2 Images for '0' and '1'.

Fig. 3 Images for '2' and '3'.

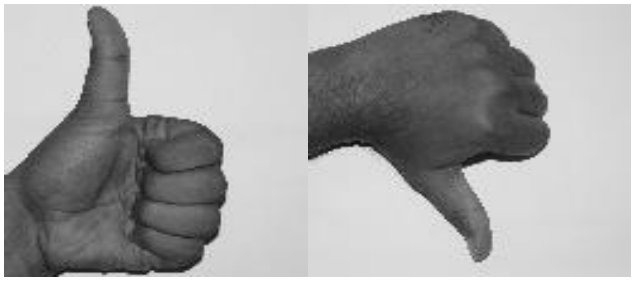

Fig. 4 Images for 'good' and 'bad'.

Fig. 4 Model extracts the feature.

The above image is a sign language of '0'. The white color wall behind is a hand, and our model extracted the main part only which was the hand.

Fig. 5 Graph for classes for Alphabets and Numbers

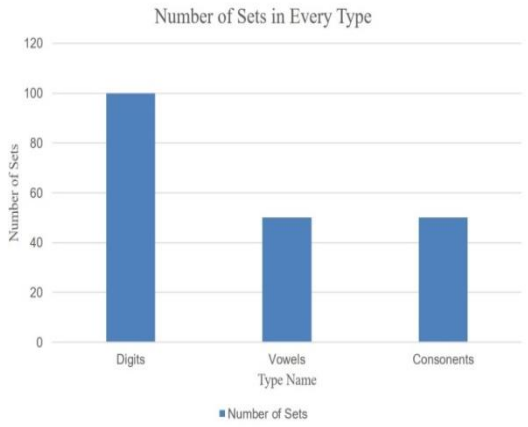

Fig. 6 Graph for the number sets, gigits, and alphabets.

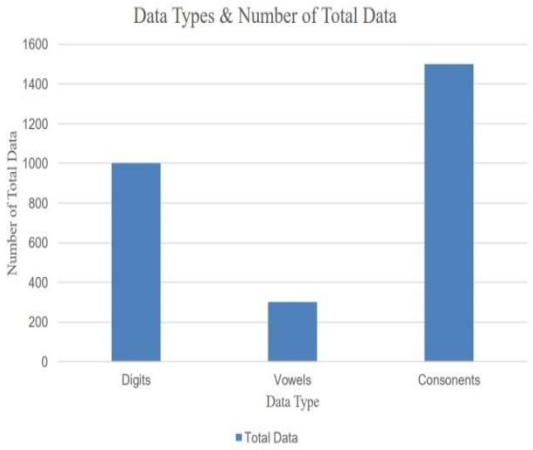

Fig. 7 Graph of total data for digits and alphabets.

IV. RESULT AND DISCUSSION

We have developed a dataset for Bangla Sign language which consists of 24168 samples (basic characters: 18745 and numerals: 5423). The proposed model generated 100% testing accuracy on digits or numeric numbers, and 97.5% testing accuracy for the alphabets of Bangla. The accuracy reached at 98.75% while both types of samples were used.

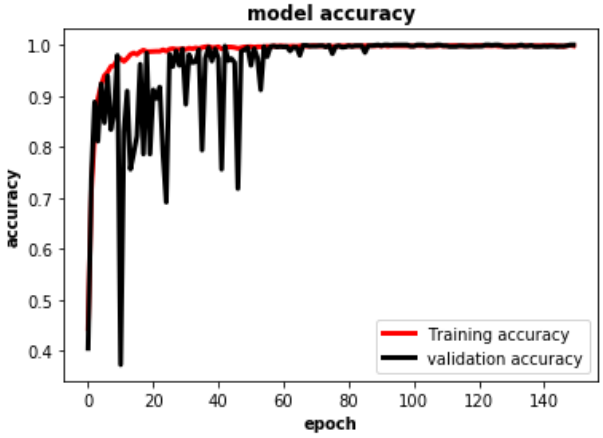

Fig. 8 Performance curve of the projected model for recognition of the digits for Bangla sign language.

With the proposed model, some extra layers like batch normalization, convolution layer's selected number, max pooling, dropout, etc. We got the highest accuracy from the digit detection. Also, the detection of alphabets was almost 98.5% which was higher than other models. The total accuracy of detection becomes slightly lower when both accuracy were used. It resulted because symbol's intra-classes added some confusion. Dataset's validity confirms classifier's performance:

Table 1. performance of the proposed model

Data Set	Training Accuracy	Test Accuracy
Characters	97.53	97.55
Digits	99.97	100
Combined	98.70	98.75

After performing a comparison and evaluation between predicted and actual values, it was that actual value shows the best and stable accuracy without any overfitting problem. Because accuracy was changing slightly after each iteration. Table1 showed the performance of the datasets. With an individual dataset, the comparison happened in real-time. From the Table 2 it is obvious that our technique is well-organized than the others.

Table 2: comparison between existing and proposed systems

Methodology	Dataset size	Accuracy
PCA	2020	76.9%
LDA	2000	88.55%
SVM	4800	97.7%
Proposed Model	5000	98.75%

Our method performed better than other as comparison. The proposed method scored the accuracy of (98.75%). The accuracy could be increased due to better hardware support. With 30FPS, it could read and detect better, showed a better accuracy. Also, some of the characters were almost the same, or identical, sometimes the model couldnot be able to detect those. But after adding some layers and more sample images for datasets could overcome the problem.

V. CONCLUSION

In this paper, we developed a model to detect Bangla Sign Language using the Convolution Neural Network (CNN). A publicly available sign language-related standard dataset was used. The dataset was split into 80:20 ratio where 80% was used for training purpose and rest of them was test purpose. CNN was able to detect and classify the signs in the images into Bangla digits and numbers. The accuracy achieved was 98.75%. The comparative result (table 2) shows that our proposed model scored the highest accuracy among other models.

In the future, we aim to make an array of images where a sign in each frame of images will be identified and classified, and sequentially concatenate them to form an expressive word or phrase or statement.

REFERENCES

[1] R. H. Abiyev, M. Arslan, and J. B. Idok, "Sign Language Translation Using Deep Convolutional Neural Networks," (in English), Ksii Transactions on Internet and Information Systems, Article vol. 14, no. 2, pp. 631-653, Feb 2020.

[2] S. Agyefi-Mensah, Z. E. Kpamma, and D. E. Hagan, "Residential adaptations as users' tacit means of communicating spatial needs in housing design A case study," Journal of Engineering Design and Technology.

[3] G. Austin, M. J. Duncan, and T. Bell, "Codesigning Parks for Increasing Park Visits and Physical Activity in a Low-Socioeconomic Community: The Active By Community Design Experience," Health Promotion Practice, Art. no. 1524839919900768.

[4] K. D. Cantin-Garside, Z. Y. Kong, S. W. White, L. Antezana, S. Kim, and M. A. Nussbaum, "Detecting and Classifying Self-injurious Behavior in Autism Spectrum Disorder Using Machine Learning Techniques," Journal of Autism and Developmental Disorders.

[5] F. Csoka, J. Polec, T. Csoka, and J. Kacur, "Recognition of Sign Language from High Resolution Images Using Adaptive Feature Extraction and Classification," (in English), International Journal of Electronics and Telecommunications, Article vol. 65, no. 2, pp. 303-308, 2019.

[6] F. Csoka, J. Polec, T. Csoka, and J. Kacur, "Recognition of Sign Language from High Resolution Images Using Adaptive Feature Extraction and Classification," (in English), International Journal of Electronics and Telecommunications, Article vol. 65, no. 2, pp. 303-308, 2019.

[7] A. Demidov, T. Syrina, and A. Tretyakov, "Development of Digital Skills and Media Education System: From the Organization of Environmental Education of Preschool Children to the ICT Competence of Teachers," Media Education-Mediaobrazovanie, no. 1, pp. 11-23, 2020.

[8] J. B. Eisengart, K. E. King, E. G. Shapiro, C. B. Whitley, and J. Muenzer, "The nature and impact of neurobehavioral symptoms in neuronopathic Hunter syndrome," Molecular Genetics and Metabolism Reports, vol. 22, Mar 2020, Art. no. 100549.

[9] S. Hossain, D. Sarma, F. Tuj-Johora, J. Bushra, S. Sen, and M. Taher, "A Belief Rule Based Expert System to Predict Student Performance under Uncertainty," in 2019 22nd International Conference on Computer and Information Technology (ICCIT), 2019, pp. 1-6.

[10] B. Johnson, C. Jobst, R. Al-Loos, W. He, and D. Cheyne, "Individual differences in motor development during early childhood: An MEG study," Developmental Science, Art. no. e12935.

[11] F. Ahmed, J. Fatema Tuj, R. J. Chakma, S. Hossain, and D. Sarma, "A Combined Belief Rule based Expert System to Predict Coronary Artery Disease," in *2020 International Conference on Inventive Computation Technologies (ICICT)*, 2020, pp. 252-257.

[12] S. Hossain, D. Sarma, F. Tuj-Johora, J. Bushra, S. Sen, and M. Taher, "A Belief Rule Based Expert System to Predict Student Performance under Uncertainty," in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, 2019, pp. 1-6.

[13] M. W. Long et al., "Evaluation of a pragmatic trial of a collaborative school-based obesity prevention intervention in a low-income urban district," Preventive Medicine, vol. 133, Apr 2020, Art. no. 106020.

[14] A. K. McCullough, M. Rodriguez, and C. E. Garber, "Quantifying Physical Activity in Young Children Using a Three-Dimensional Camera," Sensors, vol. 20, no. 4, Feb 2020, Art. no. 1141.

[15] A. A. Q. Mohammed, J. C. Lv, and M. S. Islam, "A Deep Learning-Based End-to-End Composite System for Hand Detection and Gesture Recognition," (in English), Sensors, Article vol. 19, no. 23, p. 23, Dec 2019, Art. no. 5282.

[16] K. O'Hearn, B. Larsen, J. Fedor, B. Luna, and A. Lynn, "Representational similarity analysis reveals atypical age-related changes in brain regions supporting face and car recognition in autism," Neuroimage, vol. 209, Apr 2020, Art. no. 116322.

[17] A. D. J. Pearson et al., "ACCELERATE and European Medicines Agency Paediatric Strategy Forum for medicinal product development of checkpoint inhibitors for use in combination therapy in paediatric patients," European Journal of Cancer, vol. 127, pp. 52-66, Mar 2020.

[18] S. Purkayastha, M. J. Roy, and N. Pradhan, "A demonstration for the detection of sign language," (in English), Journal of Information & Optimization Sciences, Article vol. 40, no. 8, pp. 1611-1621, 2019.

[19] S. Purkayastha, M. J. Roy, and N. Pradhan, "A demonstration for the detection of sign language," (in English), Journal of Information & Optimization Sciences, Article vol. 40, no. 8, pp. 1611-1621, 2019.

[20] T. Schack and D. Pollmann, "Promotion of development of childhood motor function Motor diagnostics and interventional concepts in kindergarten and elementary school age," Monatsschrift Kinderheilkunde, vol. 168, no. 3, pp. 215-221, Mar 2020.

[21] M. G. Sergeeva et al., "Technology Of Pedagogical Conflict Management Within An Educational InstitutioN," Revista Inclusiones, vol. 7, pp. 217-228, Apr-Jun 2020.

[22] Z. Shitu, M. M. T. Aung, T. H. T. Kamauzaman, and A. Ab Rahman, "Prevalence and characteristics of medication errors at an emergency department of a teaching hospital in Malaysia," Bmc Health Services Research, vol. 20, no. 1, Jan 2020, Art. no. 56.

[23] I. van de Kolk et al., "Study Protocol for the Evaluation of "SuperFIT", a Multicomponent Nutrition and Physical Activity Intervention Approach for Preschools and Families," International Journal of Environmental Research and Public Health, vol. 17, no. 2, Jan 2020, Art. no. 603.

[24] Z. Wahl-Alexander and C. A. Morehead, "An Observational Assessment of Physical Activity Levels and Social Behavior During Residential Summer Camp Unstructured Time," American Journal of Health Promotion, vol. 34, no. 4, pp. 387-392, May 2020, Art. no. 0890117119897191.

[25] H. Y. Yang and X. L. Curdt-Christiansen, "Conflicting linguistic identities: language choices of parents and their children in rural migrant workers' families," Current Issues in Language Planning.

[26] S. S. Shanta, S. T. Anwar, and M. R. Kabir, "Bangla sign language detection using sift and cnn," in 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2018: IEEE, pp. 1-6.

[27] S. Hossain et al. "A Critical Comparison between Distributed Database Approach and Data Warehousing Approach." International Journal of Scientific & Engineering Research, Article 5.1 (2014): 196-201.

[28] K. Noor et al., "Performance analysis of a surveillance system to detect and track vehicles using Haar cascaded classifiers and optical flow method," 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), Siem Reap, 2017, pp. 258-263.

[29] S. Hossain, D. Sarma, R. J. Chakma, W. Alam, M. M. Hoque, and I. H. Sarker, "A Rule-Based Expert System to Assess Coronary Artery Disease Under Uncertainty," in Computing Science, Communication and Security, Singapore, 2020, pp. 143-159: Springer Singapore.

[30] S. Hossain, A. Abtahee, I. Kashem, M. M. Hoque, and I. H. Sarker, "Crime Prediction Using Spatio-Temporal Data," in Computing Science, Communication and Security, Singapore, 2020, pp. 277-289: Springer Singapore.

[31] H. Alqahtani et al., "Cyber Intrusion Detection Using Machine Learning Classification Techniques," in Computing Science, Communication and Security, Singapore, 2020, pp. 121-131: Springer Singapore.