Towards Machine Learning based Bangla Sign Language Detection from Image data

by

Md Shakil Ahmed(Exam Roll: 192330)
Md Monir Hossain(Exam Roll: 192335)
Amit Azim Amit(Exam Roll: 192347)

A Researech Project Report submitted to the
Institute of Information Technology
in partial fulfillment of the requirements for the degree of
Bachelor of Science in
Information and Communication Technology

Supervisor: Dr. M. Shamim Kaiser, Professor

Institute of Information Technology
Jahangirnagar University
Savar, Dhaka-1342

November 2023

DECLARATION

We hereby declare that this thesis is based on the results found by ourselves. Materials of work found by other researcher are mentioned by reference. This thesis, neither in whole nor in part, has been previously submitted for any degree.

Md Shakil Ahmed	Md Monir Hossain	Amit Azim Amit
Roll:192330	Roll:192335	Roll:192347

CERTIFICATE

This is to certify that the thesis entitled **Towards Machine Learning based Bangla Sign Language Detection from Image data** has been prepared and submitted by **Md Shakil Ahmed**, **Md Monir Hossain** and **Amit Azim Amit** in partial fulfillment of the requirement for the degree of Bachelor of Science (honors) in Information Technology on November 9, 2023.

Prof. Dr. M. Shamim Kaiser
Supervisor

Accepted and approved in partial fulfillment of the requirement for the degree Bachelor of Science (honors) in Information Technology.

Prof. Dr. Fahima Tabassum

Chairman

Prof Dr. Mohammad Shahidul Islam

Member

Prof Dr. M. Mesbahuddin Sarker

Member

Prof. Dr. Md. Hasanul Kabir

External Member

ACKNOWLEDGEMENTS

We feel pleased to have the opportunity to express our heartfelt thanks and gratitude to those who all rendered their cooperation in making this report.

This thesis is performed under the supervision of Dr. M. Shamim Kaiser, professor, Institute of Information Technology (IIT), Jahangirnagar University, Savar, Dhaka. During the work, he has supplied us with a number of books, journals, and materials related to the present investigation. Without his help, kind support, and generous time span he has given, we could not have performed the project work successfully in due time. First and foremost, we wish to acknowledge our profound and sincere gratitude to him for his guidance, valuable suggestions, encouragement, and cordial cooperation.

We express our utmost gratitude to Dr. M. Shamim Kaiser, Director, IIT, Jahangirnagar University, Savar, Dhaka, for his valuable advice that has encouraged us to complete the work within the time frame. Moreover, we would also like to thank the other faculty members of IIT who have helped us directly or indirectly by providing their valuable support in completing this work.

We express our gratitude to all other sources from where we have found help. We are indebted to those who have helped us directly or indirectly in completing this work.

Last but not least, we would like to thank all the staff of IIT, Jahangirnagar University, and our friends who have helped us by giving their encouragement and cooperation throughout the work.

ABSTRACT

The system for detecting Bangla sign language converts it into text, enabling deaf and silent individuals to converse with regular people. Due to the lack of interaction that exists between normal people and deaf-mute persons, deaf-mute people are fairly isolated from society. According to advancements in technology, it is now feasible to record the gestures of deaf-mute individuals and use machine learning to translate them into text. It makes a bridge between deaf-mute people and normal people. The objective of our project is to increase the robustness and accuracy of sign language recognition by utilizing the powers of attention networks and capsule networks. This work introduces a machine learning based method for Bangla Sign Language Detection (BdSL) which employs attention network and capsule network. Capsule networks are a good option for sign language identification, where precise motions and postures are crucial, and their proven ability to capture hierarchical patterns in visual data. The model's ability to concentrate on important areas of interest within the sign gestures is another way that the attention networks improve performance. Our proposed model is trained on a sizable dataset of BdSL gestures, depending on both static and dynamic signs. We apply a convolutional neural network as the feature extractor that trains into a capsule network representation. The attention network is combined into a capsule network to refine the representations and allow the model to better discriminate between similar signs and handle variations in signing speed and style. This study improves communication accessibility within the BdSL community and lays the foundation for the development of similar systems for other disadvantaged sign languages. Furthermore, the combination of capsule networks and attention processes demonstrates their potential for sign language recognition, which offers valuable insights for the advancement of this field. The planned endeavor will have a major impact on the lives of persons with hearing impairments and will make a substantial addition to the field of sign language recognition.

Keywords: Machine Learning, capsule network, attention network, sign language recognition, and Bangla Sign Language.

LIST OF ABBREVIATIONS

BdSL	Bangla Sign Language Detection
CNN	Convolutional Neural Network
R-CNN	Region-Based Convolutional Neural Network
DCNN	Deep Convolutional Neural Networks
ANN	Artificial Neural Network
YOLO	You Only Look Once
CapsNet	Capsule Network
CAN	Capsule and Attention Network
BNFD	Bangladesh National Federation of the Deaf
SIFT	Invariant Feature Transform
PCA	Principal Component Analysis
LSTM	Long Short-Term Memory
MNIST	Modified National Institute of Standards and Technology database
ReLU	Rectified Linear Unit
BERT	Bidirectional Encoder Representation Transformer
ML	Machine Learning

LIST OF FIGURES

Figure

LIST OF TABLES

Table

TABLE OF CONTENTS

CHAPTER

CHAPTER I

Introduction

Bangla Sign Language (BdSL) is critical for the deaf in Bangladesh, yet current techniques are inadequate. Traditional methods have difficulty dealing with complex hand motions, facial expressions, and body movements. This study employs attention and capsule networks to better identify BdSL. Through the creation of advanced technology identification technology, it seeks to reduce the communication gap between the deaf and hearing populations in Bangladesh. Data gathering, model creation, and assessment of performance under different lighting and signer situations are all part of the research. The local influence has the potential to aid over a million deaf and hard-of-hearing people, while the international impact provides light on the state of the art in sign language recognition technology [2].

1.1 Background

For the deaf and hard-of-hearing people around the world, sign language is an essential communication tool. BdSL the national sign language of Bangladesh, is essential to enhance communication. But even with its crucial role, there aren't many automated methods available for BdSL recognition and interpretation. There is a substantial communication gap between the hearing and the deaf communities as a result of this long-lasting impairment. The main obstacle is the difficulty of sign language, which entails complex body motions, facial emotions, and hand gestures. Since traditional methods frequently fail to capture details of this visual language, they have struggled to attain high accuracy in BdSL recognition. The development of deep learning methods presents a chance to create BdSL recognition algorithms that are more reliable and accurate. The goal of this study is to increase the precision and efficacy of BdSL detection by utilizing the capabilities of attention networks and capsule networks. Attention networks allow the model to concentrate on the most related regions of the signing subject, whereas capsule networks are superior at recognizing intricate visual patterns and capturing spatial hierarchies.

1.2 Problem Statement

The fundamental difficulty addressed in this study is the absence of reliable and efficient algorithms for Bangla Sign Language (BdSL) identification [3]. The inability of traditional methods to translate BdSL's complexity has created communication difficulties for the deaf and hard of hearing in Bangladesh. Most of the studies are based on image datasets that are statics to detect sign language. They are unable to communicate successfully with the hearing community because of the lack of available BdSL recognition technologies. With the goal of making communication easier for the deaf and hard of hearing in Bangladesh, this study sets out to provide a unique solution for video datasets with dynamic properties by merging capsule networks and attention networks to improve BdSL identification [4].

1.3 Objectives

The purpose of the proposed BdSL is to develop automatic systems that can accurately recognize and interpret Bangla sign language. Deaf and hard-of-hearing persons in Bangladesh would benefit greatly from this because it would improve their ability to converse with those who are hearing.

- To perform exploratory data analysis on the selected dataset.
- To develop a ML approach for BdSL detection using a CAN.
- To evaluate the performance of the CAN model on a large dataset of BdSL gestures.

1.4 Scope of the Study

1. **Data Collection:**An exhaustive catalogue of signs and gestures in Bangladesh Sign Language (BdSL) has been compiled, with all the usual nuances and complexities accounted for.
2. **Model Development:** This paper details the design and implementation of a novel model for BdSL detection that combines attention networks and capsule networks.

3. **Performance Evaluation:** Evaluation of the model's accuracy and dependability through a battery of in-depth tests that account for a wide range of factors, such as variances in illumination, signers, and environments in which BdSL is used.

4. **User-Focused Approach:** A method that centres on the needs and preferences of the people who would be using it, in this case the deaf and hard of hearing people of Bangladesh.

1.5 Impact of the Study

1.5.1 National Impact

The lives of Bangladesh's deaf and hard people would be profoundly impacted by the creation of a novel and effective technique for detecting BdSL. More than a million people in Bangladesh are classified as deaf and hard of hearing. However, it can be difficult to find sign language interpreters in Bangladesh. This causes obstacles for deaf and hard-of-hearing people to access essential services including education, employment, and healthcare. [5].

1.5.2 Internation Impact

Beyond the national focus, this study adds to the larger subject of sign language recognition. The combination of capsule networks and attention networks used to solve the problem of BdSL identification provides a useful case study for other nations attempting to understand their own local sign languages. It has the potential to stimulate improvements in assistive technology and human-computer interaction for sign language users globally[6].

Sustainable Development Goal 4 (SDG-4) is to guarantee all people access to a good education anywhere in the world. International cooperation is essential for the successful detection of Bangla Sign Language. The deaf community in Bangladesh and around the world can benefit greatly from the creation of sign language recognition technologies. These technological advancements eliminate language barriers, allowing all students to participate in class. International cooperation can multiply this effect, making high-quality education available to more deaf people in more places. Through international Bangla Sign Language detection activities, we can realise SDG-4 and create a more just and inclusive society by expanding access to education for people of all hearing levels[7].

1.6 Report Outline

The rest of the report is structured as follows: In **Chapter II** This chapter provides a comprehensive overview of the literature and relevant work on the topic of sign language recognition, focusing on the technology and methodologies used in prior studies. **Chapter III** This chapter outlines the methodology used in this research, including Research Design, Conceptual Framework, Gantt Chart, Participants, Data Collection, and Data Analysis. **Chapter IV** This chapter contains experimental data and provides a thorough discussion of those results, including the performance and limitations of the underlying model. **Chapter V** Lastly the report's final section sums up its findings and emphasizes the significance of the suggested technique for detecting Bangla Sign Language and its larger implications for sign language identification globally.

CHAPTER II

Literature Review

In this section, it is discussed about the summary and gaps in the literature. We include a summarization of 26 research papers here. We brief all of the limitations and future works for those papers. The relevant literature review table consists of Algorithms, data sets, Evaluation metrics, Limitations, and Future Works[8]. In those studies, there are several research gaps where most of the works are based on images in English and other languages. In the research field, there is not enough work or studies with Bangla sign language based on video datasets.

2.1 Summary of the Relevant Literature

Oishee Bintey Hoque and Mohammad Imrul Jubair developed real-time signs from images using R-CNN based on BdSLImset (Bangladeshi Sign Language Image Dataset) in 2018. They get a recognition time of 90.03ms and a loss of 0.07538 [9]. In the titled Bangladeshi Hand Sign Language Recognition from Video the author using algorithms LBP and SVM got an accuracy of 94.26% for words and 94.49% for sentences [10]. For the Recognition of BdSL using a Convolutional Neural Network, the accuracy is 99.80% using the BdSL dataset of 30916 samples and based on the CNN model [11]. Using YOLOv4 as the object detection model, the authors get an accuracy of 99.95% based on the dataset of 12.5k images of 49 different signs [12]. Deep Learning-based Bangla Sign Language Detection with an Edge Device achieved an accuracy of 94.91% using a custom dataset and their used model are various deep learning techniques, Detectron2, EfficientDet, and YOLOv7 in 2023 [13]. Their recognition, however, was not in real-time. The authors also presented a strategy for recognizing BdSL letters and digits in real-time using a fuzzy-logic-based model and grid-pattern analysis in [14]. The authors of [15] described a real-time Bengali and Chinese number sign recognition system based on contour matching. The system was trained and evaluated using a total of 2000 contour templates from 10 signers for both Bengali and Chinese numerical signs, achieving recognition accuracy of 95.80% and 95.90% with a computational cost of 8.023 milliseconds per frame. Alvaro Budria, and Laia Tarres use the how2sign dataset based on the model Neural Architecture, LSTM and they get 70%. The authors of [16] used Convolution Neural Networks to create an Arabic Sign Language Recognition and Speech Generation system. They utilized the Google Translator API to convert from hand sign to letter, and then gTTs were used to generate voice. There have been several studies on BdSL detection, as well as other Sign Languages. The authors of [17] created a Bangla Sign Language recognition system in which they translated training data from RGB to HSV color space and then extracted features using the Scale Invariant Feature Transform (SIFT) Algorithm before feeding to the model. The implementation was completed in order to identify 38 Bangla signals. Another study [18] employed the Support Vector Machine Algorithm to recognize Bangla Hand Sign Language. They also transformed photos from RGB to HSV color space for data preparation, but unlike earlier researchers, they identified features using the Principal Component Analysis (PCA) technique to minimize dimensionality before feeding the data to the model.

Table 2.1: Literature Review Part-I

Reference	**Algorithms**	**Datasets**	**Limitations**
[9]	CNN, Faster R-CNN	BdSLImset (Bangladeshi Sign Language Image Dataset)	The model faced is with facial features and skin tones.
[10]	LBP, SVM		Not For work Video
[3]	Capsule Network	The Ishara-Lipi dataset	Only Detect Bangla Digit
[11]	CNN	BdSL dataset of 30916 samples	Troublesome for most of the people who are not acquainted with the BdSL to communicate without an interpreter.
[16]	CNN	Own Dataset	Even with a small sample size, the error rate was quite low, proving the method's reliability.
[19]	DCNN	The dataset comprises 37 hand signs (total 1147 images),	Work with one-handed signs, photos captured using a mobile camera up to 13 mp, anther dataset has few samples as input
[12]	YOLO v4 as the object detection model	Used dataset consisting of 12.5k images of 49 different signs	Contains signs for only 36 alphabets, Bengali punctuation "l", there are still no signs available for other punctuations.
[13]	with various deep learning techniques, Detectron2, EfficientDet, and YOLOv7	Paper employs two datasets: Okkhornama (Talukder et al., 2021) and custom dataset with 46 BSL characters.	N/A
[20]	CNN	1500 images from 10 different users	Only 13 of the 49 Bengali alphabets are not represented in sign language, inability to generate some commonly used words “বৃদ্ধ”, “ঋিষ”, “ঊষা”, ”বাঙাল”.

Table 2.2: Literature Review Part-II

Reference	**Algorithms**	**Datasets**	**Limitations**
[21]	YOLOv4	Dataset: 12.5k images of 49 signs, comprising 10 numerals, 36 BdSL characters, and 3 proposed signs (characters, space, end sentence).	N/A
[22]	artificial neural network (ANN)	BSL image dataset	To create a floating-point number, the decimal point must be present.
[5]	CNN as a Deep Learning method	A digital camera captured 1005 samples of the 36 signed Bangla letters and the accompanying depth data.(created by ourselves)	N/A
[4]	MediaPipe Holistic and LSTM	own collection of data	N/A
[23]	MobileNetV2, conditional deep convolutional generative adversarial network, CNN, Computer vision	used the dataset provided by Rafi et al. on kaggle.com	In each training, the model restored the parameters for which we got the lowest validation loss.
[2]	CNN	N/A	There are 1100 photos in all, labeled with one of 11 different hand gestures. They hope to increase it.
[8]	CNN, RNN,Computer Vision	American Sign Language Dataset.	N/A
[24]	Neural Architecture, LSTM	How2Sign dataset.	

Table 2.3: Literature Review Part-III

Reference	Algorithms	Datasets	Limitations
[25]	CNN	used Open computer vision(OpenCV) library in	data set accessibility, picking a good picture filter for a CNN to use to extract features, etc.
[26]	capsule networks, LeNet	sign language MNIST	N/A
[6]	Capsule networks	Used four public sign language datasets, i.e., NMFs-CSL, SLR500, WLASL, and MSASL.	N/A
[7]	CNN, RNN	RWTH-PHOENIX-Weather 2014T	Further enrich the RWTH-BOSTON-400 annotations, and will open up the path to multiple stream processing
[27]	Convolutional Neural Network	collected 25 sign language gestures, 100 training pictures for each gesture, and trained by CNN.	At present, our system can only translate separate words.
[28]	(CNN)	Own dataset	N/A

2.2 Gaps in the Literature

These model's facial characteristics and skin tone were one of its many challenges. Video Only Detect Bangla Digits, Not Safe for Work, It is troublesome for most of the people who are not acquainted with the BdSL to communicate without an interpreter. We were able to prove the method's robustness from prior work while working with a smaller data set than was originally requested. Sign with one hand as you work. a picture taken using a cell phone camera that's up to 13 megapixels in resolution. There are a few input examples in the dataset. Only 36 alphabets' worth of symbols are included[12]. The "I" is the only sign available for the Bengali punctuation system at the moment. There are 49 letters in the Bengali alphabet, but only 36 corresponding signals in the sign language. A lack of capacity to come up with the terms বৃদ্ধ, ঋিষ, ঊষা, বাঙিাল [21]. Since there is no decimal point symbol, floating-point numbers cannot be formed. Although there are 49 alphabets in the

Bengali Language, sign language has signed for only 36 alphabets. The lack of a decimal point prevents the creation of floating-point numbers. In each training, the model restored the parameters for which we got the lowest validation loss. Only 11 hand gesture labels were utilized in this collection, with 100 photos for each label. They're aiming to boost the number of photographs in their current dataset. A lack of data was the first challenge we encountered. Second, we needed to decide on a filter to apply to our photos to extract the right features, and only then could we use that image as input for the CNN model. At present, our system can only translate separate words.

CHAPTER III

Methodology

3.1 Research Design

In figure 3.1 represents a camera image is taken for Bangla Sign Language recognition. To improve the image's quality and separate the sign language gestures, preprocessing is performed. A training dataset is built from the processed data. To decipher the indications, a Capsule Network (CapsNet) classifier is educated on this data set. The effectiveness of the network is measured, and the results show that it can accurately recognize Bangla Sign Language motions, which might be used to help the deaf and hearing-impaired communicate[29].

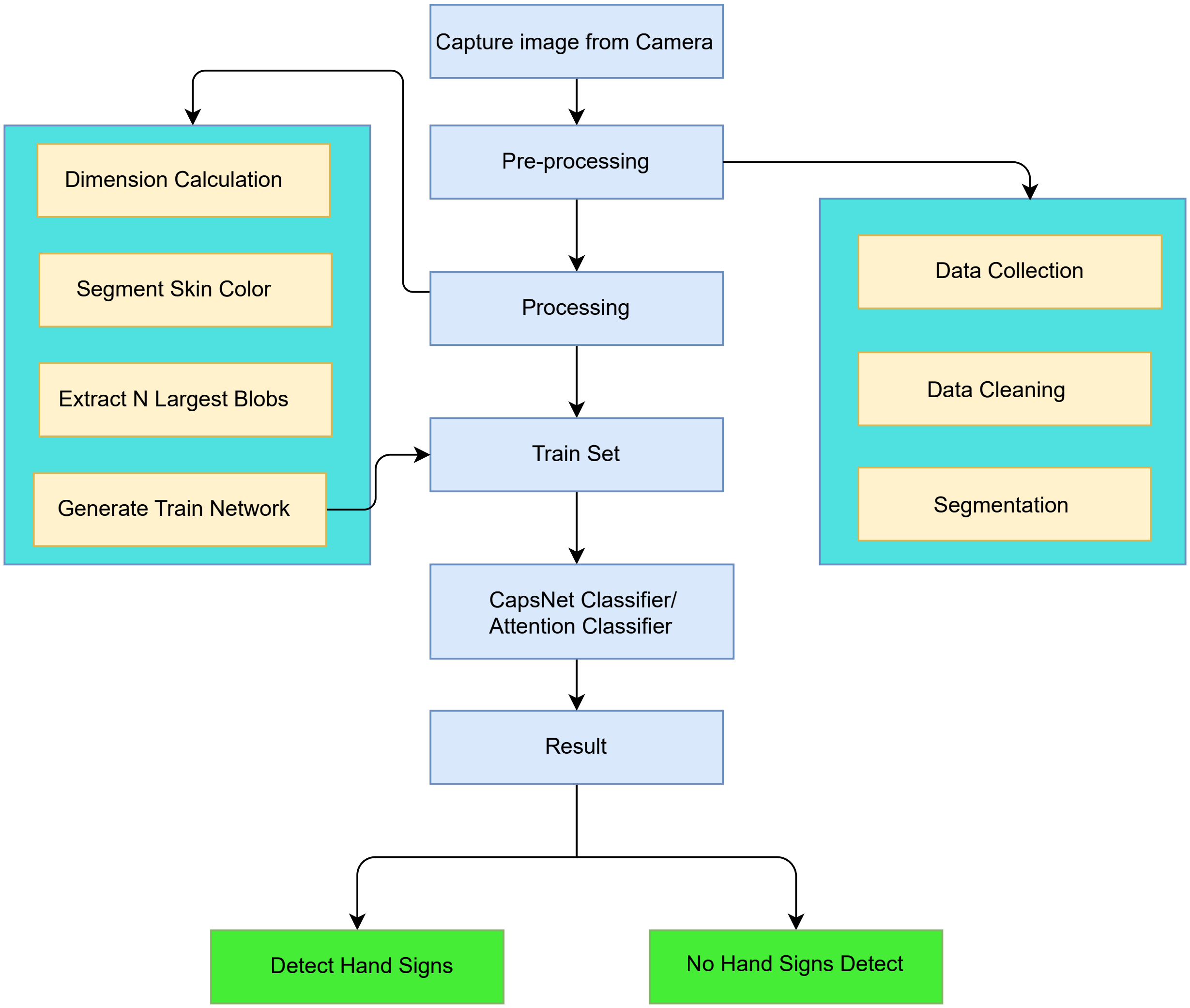

Figure 3.1: Proposed Model Diagram

3.2 Conceptual Framework

Using a convolutional neural network (CNN) to extract features from sign language videos, followed by a capsule network and an attention network to classify the extracted features, is the conceptual framework for Bangla sign language detection using the capsule network and attention network. While the attention network zeroes down on the most important aspects for classification, the capsule network derives high-level features from the CNN data.

3.2.1 Capsule Network

Machine learning systems like the artificial neural network (ANN) variant known as a capsule neural network (CapsNet) can be put to use modeling hierarchical structures[30].

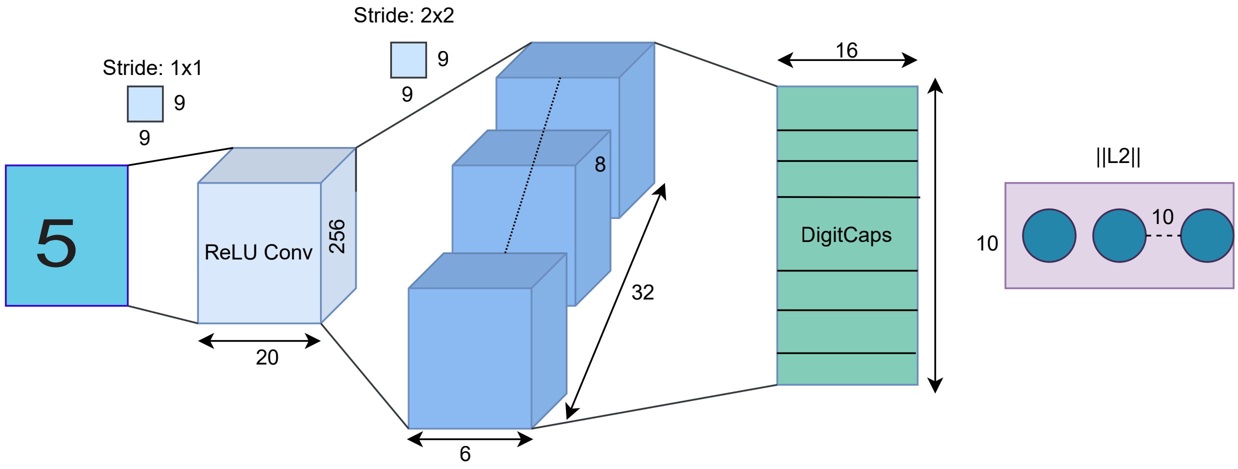

Figure 3.2: Architecture of CapsNet

In Figure 3.2, we see the first iteration of the CapsNet architecture, which yields similar results to a deep convolution network. Each class instance is represented by the length of its activation vector in the DigitCaps layer, which is then used to compute the classification loss[31].

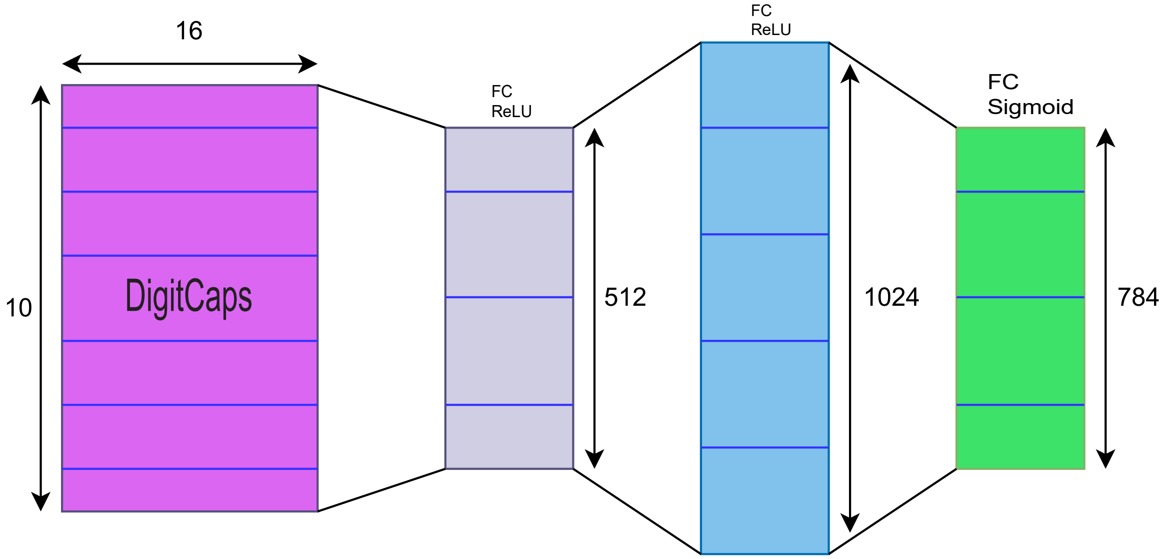

Figure 3.3: Building a decoding framework from the DigitCaps layer

DigitCaps layer decoding architecture is shown in Figure 3.3. There are two complete connection layers that ReLU and tanh use to regulate the flow of DigitCaps. Training involves minimizing the Euclidean distance between pictures and the sigmoid layer's output. Training reconstruction using the proper label as the intent[31].

3.2.2 Attention Network

Machine learning systems like the artificial neural network (ANN) variant known as a capsule neural network (CapsNet) can be put to use modeling hierarchical structures.

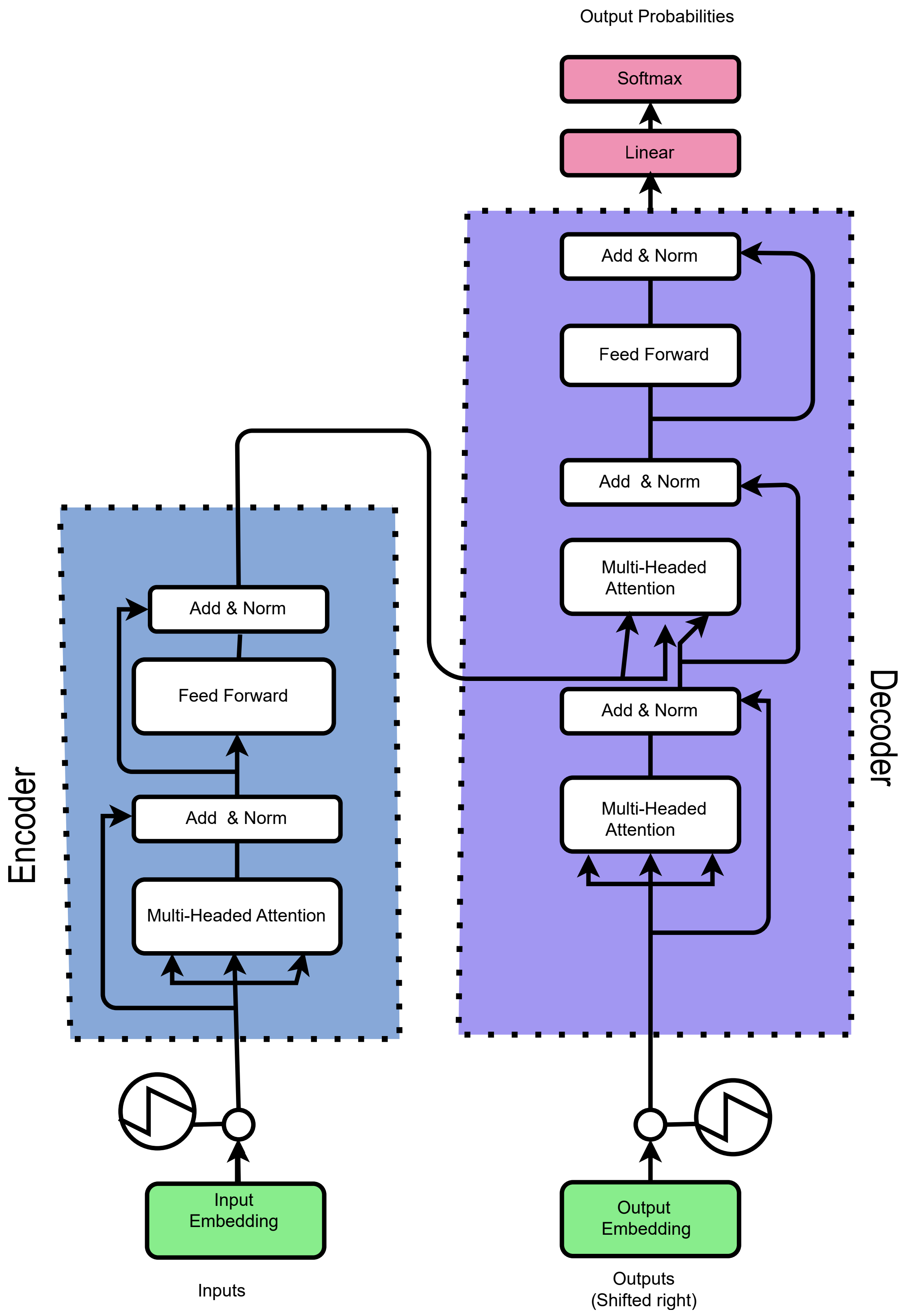

Figure 3.4: Step of Attention Network

The Bidirectional Encoder Representation Transformer (BERT) is a pre-trained model that transforms words into vectors using a multi-headed attention-based encoder-decoder. Predicting a legal ruling based on a mountain of case documentation is a breeze with this method.

The procedure of BERT architecture is shown in Fig. 3.4 [1]

1. First, the input is sent to a word embedding layer, where each word is given a vector representation, and a lookup table is built.

2. Since the encoder of the transformer lacks positional information, it is instead conveyed into embeddings. Every even and odd index uses the sine and cosine functions, respectively, for positional encoding.

3. All of the sequence data is stored in the encoder layer. There are two distinct groupings within it. Diverse viewpoints and eventually an interconnected web. The multi-headed attention model employs a self-attention mechanism, connecting each input word to those that came before and after it. The query, key, and value components serve as the basis for an autonomous attention system.

4. To get a score matrix, we just execute a dot product of Query and the key value. This provides a hint as to how seriously to take a sentence's every word. Greater the score, the more crucial those terms are. Therefore, questions become keys.

5. A Softmax is used to assign importance weights between 0 and 1 to the scaled score. When this is done, higher scores are raised to even higher levels, and vice versa.

6. Multiply the result of Softmax with the value vector to get the output vector.

7. The decoder layer, like the encoder layer, consists of a feedforward layer and two multi-target attention layers. A linear layer is added on top to serve as a classifier, and then the word.

3.2.3 Caps-Attention

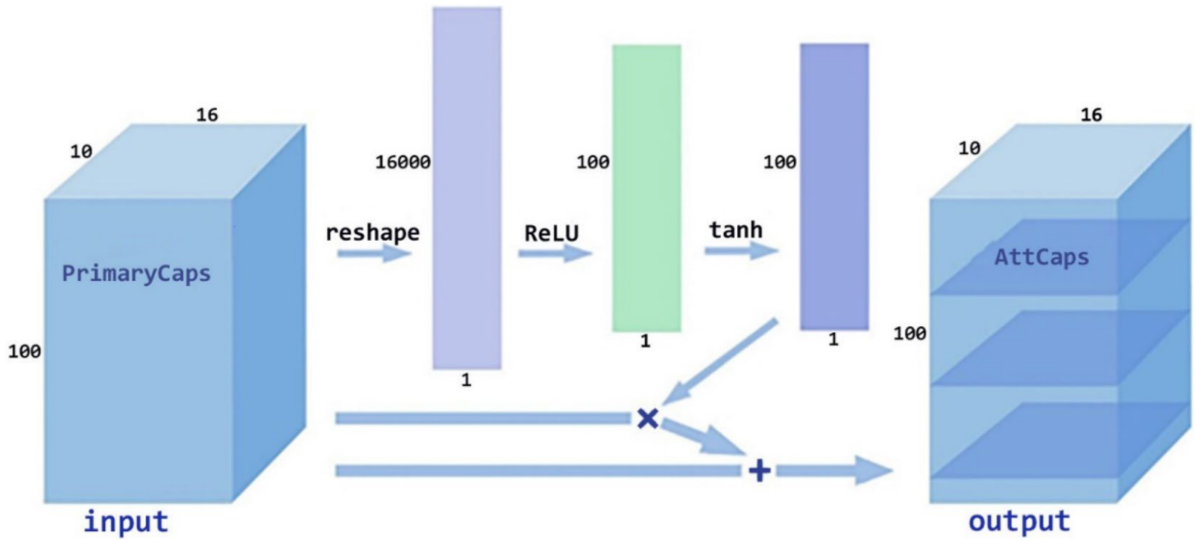

Figure 3.5: Architecture Design of Caps-Attention [1]

Figure 3.5 illustrates the concept of Caps-Attention. After transforming PrimaryCaps into a vector, sending it through a fully connected neural network with ReLU and tanh activation functions, and finally multiplying and adding it to itself, we get at AttCaps.

3.3 Gantt Chart

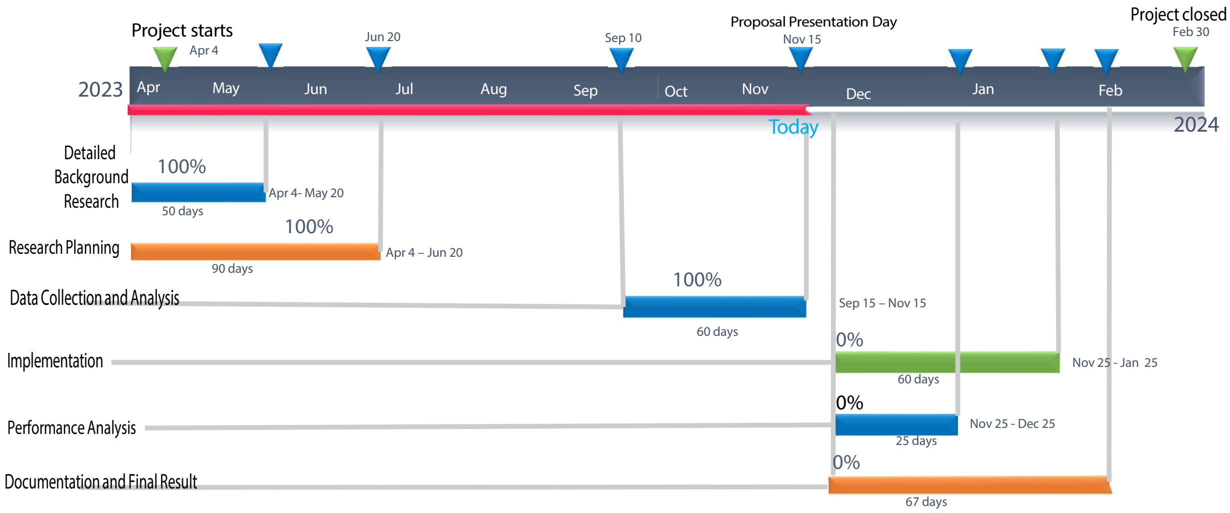

Figure 3.6: Project Implementation Plan

Figure 3.6 displays the thesis timetable, how long research planning, Data gathering analysis, implementation, Performance analysis, Documentation, and findings. We were able to conclude our study in February of this year, so we know that some aspects of research preparation take significantly longer than others.

3.4 Participants

Fluent users of Bangla sign language are encouraged to participate in our research project on Bangla sign language detection. Auditory participants who are fluent in Bangla sign language are possible. The study's scope and methodology will determine how many people will need to take part [24]. Here are some participants for recruiting in our study of Bangla sign language detection:

- **Deaf and hard-of-hearing organizations:** The people who have deaf-of-hearing problems in the organization are willing to participate in our research work.
- **Sign language workshops and classes:** We can arrange workshops for deaf and hard of hearing impairment students and collect data from the workshops.
- **Bangladesh National Federation of the Deaf (BNFD):** We have to collect our data from BNFD[22].
- **Sign Language experts:** Experts in sign language should be consulted to guarantee that the recorded motions are accurate and consistent.

3.5 Data Collection

All the data for research are being collected from the Deaf and hard-of-hearing organizations, Sign language workshops and classes, and the Bangladesh National Federation of the Deaf (BNFD). Participants are deaf and hard-of-hearing individuals. Interviews, video recordings, and other methods may be used for this purpose. We used a camera, microphone, smartphone, and other relevant equipment for collecting data [2].

3.6 Data Analysis

Data analysis is an essential part of Bangla Sign Language Detection for the development of an accurate sign language system[6].

3.6.1 Preprocess and cleaning:

A first processing step is performed on the raw data to guarantee its quality and consistency. Image normalization, noise reduction, and the elimination of outliers and erroneous data may be required for this step.

3.6.2 Feature Extraction:

To create a suitable representation of the hand motions, pertinent characteristics are retrieved from the preprocessed data. Hand position, orientation, finger position, and motion patterns are all shared characteristics [25].

3.6.3 Feature Selection:

A selection of the most useful and discriminating features is picked to minimize dimensionality and enhance model performance. Methods like principal component analysis (PCA) and feature significance measurements might be used for this purpose.

3.6.4 Model Training and Evaluation:

The chosen features are used to train a machine learning model, such as a Machine learning(ML) and convolutional Neural Network (CNN). Metrics including accuracy, precision, recall, and F1-score are used to evaluate the models.

3.6.5 Model Optimization and Hyperparameter Tuning:

To boost performance and increase generalizability to new data, trained models' hyperparameters like learning rate, kernel size, and network architecture are adjusted [26].

3.6.6 Error Analysis and Interpretation:

Analyzing the models' mistakes helps find gaps in the data or places where the models are incorrect. This aids in enhancing the overall performance and endurance of the system.

CHAPTER IV

Conclusion and Future Works

The purpose of this study is to increase text conversion accuracy in Bangla sign language by combining video sign language datasets with the Attention and Capsule network (CAN model). The majority of past research has focused on picture or static information, while this effort intends to improve sign language interpretation for video data. Enhancing the understanding of spatial hierarchies and temporal dynamics in sign language communication, notably in identifying complicated motions, hand movements, and sequential expressions, is one of the theoretical implications. On a practical level, the integration of these networks has the potential to improve accessibility for the Bangla sign language community, which will assist the deaf and hard of hearing. It has the potential to lead to real-time interpretation tools, quick connection via video platforms, and cultural diversity via offering learning resources. The system might potentially have larger uses across other sign languages, broadening its worldwide reach. Future research will concentrate on enhancing model accuracy, establishing real-time translation and interpretation, and integrating accessible features into smart and wearable devices, to eventually provide a holistic accessibility solution[23].

4.1 Summary of the Main Findings

In this study, we want to get text data from using video sign language datasets using the Attention and Capsule network. There are many limited works from using video datasets. Most of the research worked for image or static datasets. We expect more accurate text conversion from video datasets using the CAN model[1].

4.2 Restatement of the Theoretical and Practical Implications

The theoretical implications of merging capsule networks and attention processes for Bangla sign language identification in video datasets concentrate on improving sign language interpretation knowledge of spatial hierarchies and temporal dynamics. These implications imply that knowledge of the complicated motions, hand movements, and sequential expressions inherent in sign language communication is progressing. The practical implications and the integration of these networks offer great promise for enhancing accessibility for the deaf and hard-of-hearing communities who use the Bangla sign language. This advancement could lead to the development of real-time interpretation tools, allowing for instant communication via video platforms. Furthermore, it may contribute to cultural diversity by offering learning resources and educational venues to empower individuals in the Bangla sign language community. There is also the possibility of larger uses across multiple sign languages, increasing its worldwide influence[29].

4.3 Suggestions for Future Research

The use of the Capsule network and attention network for Bangla sign language identification based on video datasets has the expectation to significantly advance accessibility technologies. For future work, there is an opportunity to enhance model accuracy. Besides, we will develop real-time translation and interpretation. This system can translate BdSL into speech. We want to conclude this system integrating accessibility features in smart and wearable in the future[21].

References

[1] "Attention network." https://medium.com/@geetkal67/attention-networks-a-simple-way-to-understand-self-attention-f5fb363c736d.

[2] B. Garcia and S. A. Viesca, "Real-time american sign language recognition with convolutional neural networks," *Convolutional Neural Networks for Visual Recognition*, vol. 2, no. 225-232, p. 8, 2016.

[3] T. Hossain, F. S. Shishir, and F. M. Shah, "A novel approach to classify bangla sign digits using capsule network," in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, pp. 1–6, IEEE, 2019.

[4] M. W. Foysol, S. E. A. Sajal, and M. J. Alam, "Vision-based real time bangla sign language recognition system using mediapipe holistic and lstm," in *2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, pp. 19–24, IEEE, 2023.

[5] N. Hassan, "Bangla sign language gesture recognition system: Using cnn model," *ScienceOpen Preprints*, 2022.

[6] M. Bilgin and K. Mutludoğan, "American sign language character recognition with capsule networks," in *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pp. 1–6, IEEE, 2019.

[7] W. Zhao, H. Hu, W. Zhou, J. Shi, and H. Li, "Best: Bert pre-training for sign language recognition with coupling tokenization," *arXiv preprint arXiv:2302.05075*, 2023.

[8] 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, IEEE, 2018.

[9] O. B. Hoque, M. I. Jubair, M. S. Islam, A.-F. Akash, and A. S. Paulson, "Real time bangladeshi sign language detection using faster r-cnn," in *2018 international conference on innovation in engineering and technology (ICIET)*, pp. 1–6, IEEE, 2018.

[10] U. Santa, F. Tazreen, and S. A. Chowdhury, "Bangladeshi hand sign language recognition from video," in *2017 20th International Conference of Computer and Information Technology (ICCIT)*, pp. 1–4, IEEE, 2017.

[11] R. B. Rafiq, S. A. Hakim, and T. Tabashum, "Real-time vision-based bangla sign language detection using convolutional neural network," in *2021 International Conference on Advances in Computing and Communications (ICACC)*, pp. 1–5, IEEE, 2021.

[12] D. Talukder and F. Jahara, "Real-time bangla sign language detection with sentence and speech generation," in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, pp. 1–6, IEEE, 2020.

[13] S. Siddique, S. Islam, E. E. Neon, T. Sabbir, I. T. Naheen, and R. Khan, "Deep learning-based bangla sign language detection with an edge device," *Intelligent Systems with Applications*, vol. 18, p. 200224, 2023.

[14] M. A. Rahaman, M. Jasim, M. H. Ali, T. Zhang, and M. Hasanuzzaman, "A real-time hand-signs segmentation and classification system using fuzzy rule based rgb model and grid-pattern analysis.," *Frontiers Comput. Sci.*, vol. 12, no. 6, pp. 1258–1260, 2018.

[15] M. A. Rahaman, M. Jasim, T. Zhang, M. H. Ali, and M. Hasanuzzaman, "Real-time bengali and chinese numeral signs recognition using contour matching," in *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 1215–1220, IEEE, 2015.

[16] F. Yasir, P. Prasad, A. Alsadoon, A. Elchouemi, and S. Sreedharan, "Bangla sign language recognition using convolutional neural network," in *2017 international conference on intelligent computing, instrumentation and control technologies (ICICICT)*, pp. 49–53, IEEE, 2017.

[17] 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)*, pp. 1–6, IEEE, 2018.

[18] M. A. Uddin and S. A. Chowdhury, "Hand sign language recognition for bangla alphabet using support vector machine," in *2016 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, pp. 1–4, IEEE, 2016.

[19] 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, IEEE, 2018.

[20] R. B. Rafiq, S. A. Hakim, and T. Tabashum, "Real-time vision-based bangla sign language detection using convolutional neural network," in *2021 International Conference on Advances in Computing and Communications (ICACC)*, pp. 1–5, IEEE, 2021.

[21] D. Talukder and F. Jahara, "Real-time bangla sign language detection with sentence and speech generation," in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, pp. 1–6, IEEE, 2020.

[22] S. T. Ahmed and M. Akhand, "Bangladeshi sign language recognition using fingertip position," in *2016 International conference on medical engineering, health informatics and technology (MediTec)*, pp. 1–5, IEEE, 2016.

[23] A. Al Rafi, R. Hassan, M. Rabiul Islam, and M. Nahiduzzaman, "Real-time lightweight bangla sign language recognition model using pre-trained mobilenetv2 and conditional dcgan," in *Proceedings of International Conference on Information and Communication Technology for Development: ICICTD 2022*, pp. 263–276, Springer, 2023.

[24] M. M. Rahman, M. S. Islam, M. H. Rahman, R. Sassi, M. W. Rivolta, and M. Aktaruzzaman, "A new benchmark on american sign language recognition using convolutional neural network," in *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*, pp. 1–6, IEEE, 2019.

[25] A. Budria, L. Tarres, G. I. Gallego, F. Moreno-Noguer, J. Torres, and X. Giro-i Nieto, "Topic detection in continuous sign language videos," *arXiv preprint arXiv:2209.02402*, 2022.

[26] A. Kumar, A. Gupta, B. Chaurasia, and C. Mishra, “Sign language to text conversion using cnn (for deaf and mute people),”

[27] P. Dreuw, C. Neidle, V. Athitsos, S. Sclaroff, and H. Ney, “Benchmark databases for video-based automatic sign language recognition.,” in *LREC*, 2008.

[28] N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, “Neural sign language translation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7784–7793, 2018.

[29] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, “Sign language recognition using convolutional neural networks,” in *Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13*, pp. 572–578, Springer, 2015.

[30] “Capsule network.” https://medium.com/ai/theory-practice-business/understanding-hintons-capsule-networks-part-i-intuition-b4b559d1159b.

[31] W. Huang and F. Zhou, “Da-capsnet: dual attention mechanism capsule network,” *Scientific Reports*, vol. 10, no. 1, p. 11383, 2020.