International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Assessing Google's Teachable Machine by Deploying American Sign Language Detection System

Sanjeevkumar B[1], Varun S P[2], Yokesh Babu S[3]

[1,2]B. Tech Student, School of Computer Science and Engineering, Vellore Institute of Technology, Tamil Nadu

[3]Associate Professor, Department of IOT, School of Computer Science and Engineering, Vellore Institute of Technology, Tamil Nadu

Abstract

Sign language recognition and detection using the current technologies has become one of the focused areas of research which aims to help with communication capability and accessibility for the people of the deaf and hearing issues related community. Over the year with the improvement in the object detection space, A technology by Google 'Teachable Machine' which is a web-based tool that allows users to custom train machine learning models with the user's available data, thereby acting as an automatic ML modeler for producing data models with ease. This paper proposes a method to evaluate the Teachable Machine's effectiveness by utilizing the American Sign Language (ASL) to make the model recognize the signs once trained. The project involves data collection, image preprocessing, model training using the Teachable Machine's and testing. The results demonstrated an overall accuracy in predicting ASL letters, though the model is struggling to predict certain letters with similar symbols. The paper also proposes the real-life applicability of such systems in various domains, strengths, and limitations of teachable machines with future scope for improvements in accuracy and real-time processing.

Keywords: Sign language recognition, Object detection, Teachable Machine, Data models, Evaluate, Image preprocessing, Model training.

1. INTRODUCTION

Sign language is a type of language involving sets of actions and signs that are used by deaf and hard- of-hearing communities to converse worldwide. Despite its relevance and importance there are lot of barriers related to accessibility, including limited availability of sign language interpreters and lack of knowledge about the different signs among the common or general population. Thus, sign language systems have potential to reduce this gap between automatically by interpreting and translating the sign language gestures and into common spoken language. These types of tools and technologies can help enhance accessibility in various domains from healthcare, education etc.

Google's Teachable Machine is a web-based tool that helps and tries to make machine learning more adaptable and user-friendly. It is a tool that allows users to train their own custom learning algorithms.

The tool also supports various input data forms including images audio and sensor data. Here in our paper, we will shed light to this interface of the Google's new automatic Machine learning model with the by training it with the ASL letter and try to test the same and identify it prediction capabilities

The existing methods from the simplest ones of using TensorFlow to identify signs [1] to modifying the LSTM [2], RF Sensors [3], [5] , or using pattern matching techniques to recognize the sign [7] with Edge Detection and Cross Correlation methods[7] or using SVM classifier[9] and various other Neural Networks like such as 3-D CNN [9] and RNN [11] all require complex Machine Learning knowledge and modelling apart form which the computational requirements also go higher when trying to work and process these models. Thereby we can see that there is a requirement for an efficient automatic modeler that can perform the task of producing ML model and provided platform for doing the model training task easily. Though there is need we should also assess the working of these automatic ML modelers by deploying a task to them.

This paper offers some of the key contributions for assessing the Teachable Machine (Automatic ML modeler) by deploying an American Sign Language Detection to it and assess using probabilistic test. Also, we prove that the Teachable machine is effective in creating smaller models but might not be that effective for larger Data . Through this we also provide an alternative to model small sign detection systems where one does not need to know complex machine learning algorithms. We also evaluate the accuracy of the automatic ML model's performance based on the deployed ASL detection task to the model and provide the models merits and demertis.

2. Related work

Bhatt et al. introduce a new approach to real-time American Sign Language detection in their paper done in 2024 [1]. The significance of this work lies in the rapidly evolving field of sign language recognition, primarily driven by the advance in deep learning solutions. While these technologies initially attracted wide interest from researchers working with the Indian Sign Language or ISL, the current represents an area wherein the examined subject is ASL, which allows the proven results to have more direct influence on the way hearing people perceive the deaf community.The modified LSTM architecture for continuous sign language recognition using leap motion senor by Mittal et.al [2], differs many of the hidden model, artificial neural network and depth camera which inly deals with single sign gesture because they consider long sequences of sign they make this by 3D extraction of hand/finger positions with CNN for spatial feature extraction and resetting reset gates in long short- term memory (LSTM). When evaluated on a dataset containing 942 sentences in Indian Sign Language (ISL) where each sentence consists of 35 words separated by pauses. The model achieves 72.3% accuracy on sentences and 89.5% accuracy on isolated words which is better than traditional LSTMs that only recognize one word at a time. Using RF sensors as suggested [3], [4], address the issues such as wearable gloves, posing privacy concerns and mobility limitations to deaf users. The method proposed by Rahman et a l. [3] uses RF sensing for 20ASL signs, Sequential classification. The research demonstrates an accuracy level in asl word-level classification (91.3%), sequential recognition (92.3%) and trigger recognition rate (93%).similarly using RF sensors Gurbuz et al.[4] in his paper proposes to focus the use of RF sensing in distinguishing native ASL siging from imitation. The machine learning analysis achieves a 72.5% accuracy in ASL sign Classification. They underscore the importance of validating ML algorithms on native ASL data to ensure robust performance and broader accessibility for the deaf community. . In another project Shi et. al [5] works on fingerspelling recognition in "the wild" with dataset of 7,000 fingerspelling sequences collected from video streams in online space. Their dataset had issues from lighting, background disturbances and camera angles with presence of 168 different singers. As a result, they established the world letters recognition accuracy of about 42% messes up the performance of this task than previous studies .The project done by Deep et.al (2022) [6] presents a method to capture the landmarks of hand using Media Pipe and storing it in a NumPy array and later using Keras and TensorFlow for detection and recognition.

Apart from this method some of the previous and traditional image processing methods were also convincing such as Joshi et. al explained in a research paper how they developed a system that translates American Sign Language (ASL) into English text in real-time by using image processing and pattern matching techniques. [7] This method works with any computer equipped with an ordinary web camera. The images are captured from the webcam and then the team creates databases of ASL hand gestures for individual alphabet characters as well as words or phrases Their model of automated ASL translation possible without special equipment in their prototype system achieved 94.23% accuracy on recognizing individual alphabet characters while basic words/phrases recorded 92%. Yet they had issues with background noise and needed a bigger dataset. In another study by Jiang et. al [8] provides a system in view of constant vision for recognizing fixed hand motions from American Sign Language using a USB camera. It is stated that without requiring special hardware skin color detection in the HSV color space is often used for hand segmentation; however, this model also still poses challenges when dealing with complex and noisy backgrounds and lighting conditions but they expand upon these methods by suggesting adaptive skin colour segmentation, noise removal techniques, reorientation of hand gestures, PCA based feature extraction as well as SVM classifier was able to identify five ASL hand shapes with 99% accuracy. In a new paper by Sharma et al [9] which aims at enabling effective communication between the hearing and deaf communities remains a significant challenge. Focusing on American Sign Language(ASL) recognition using 3D Convolutional Neural Networks (3D CNNs) for dynamic gesture recognition and then applying it to Boston ASL Lexicon Video Dataset (LVD), with over 3,300 signed English words across 6 signers. Using a 70-30 split for training and testing, the model achieved improved precision of 3.7% higher, recall 4.3% and f-measure (3.9%). Abdulwahab et.al [10], In his paper proposes a Deep Learning approach for recognizing static ASL gestures, involving resizing ASL binary images and detecting hand boundaries. It also classifies 24 static ASL alphabet characters using Convolutional Neural Networks with an accuracy of 99.3% but they are confined to only his dataset. Employing the use of Recurrent Neural network (RNN) by lee et. al [11] who proposed a real time sign recognition system where he used LSTM-RNN and k-NN to achieve the recognition and address the need for interactive and engaging tool. The proposed model achieves an accuracy of 99.44% and a 5-ffold cross-validation accuracy of 91.82% which indicate the potential for facilitating effective ASL learning experiences. In a traditional approach by Rahib et.al
[12] in this paper proposes a object detection using Single Shot Multi Box Detection (SSD) and an inception v3 plus Support Vector Machine(SVM) which projected a 99.9% accuracy but involved a lot of complication in developing the model which we will eliminate by using the Teachable machine which will carry out the entire model preparation work. A slightly different domain where Hamid R et. al [13], introduces a large-scale, real-life sign language dataset with over 25,000 annotated videos and 1000 signs from over 200 signers. It also proposes the use of I3D, a powerful architecture known from video classifications, for sign language recognition though it outperforms it still has a complex purpose of model preparation and evaluation. Shi B et. al [14] in his paper introduces the largest fingerspelling dataset which comprises of 7,304 sequences from online videos they report a letter accuracy of 41.9% on this "in the wild" data using a hand detector and CTC-based sequence model. In contrast, prior work on more controlled studio data achieved under 10% letter error rates. Sanmitra et. al (2021) [15] in her paper uses

SSD ML as proposed earlier [12] where the model is trained for 20 images to achieve a high accuracy, but the paper finally describes a model with accuracy of only 85% which surely has scope of improvement. In a method proposed by Fang B et. al [16] based on deep learning, for translation of American Sign Language (ASL) to both words and sentences without direct interference, capturing signs using infrared sensors. A hierarchical bidirectional RNN with a CTC-based probabilistic framework is utilized and is noted to reach 94.5% word-level accuracy and to obtain 8.2%-word error rate on a dataset of 7,306 samples from 11 participants with 56 words and 100 sentences. These technologies bode well in attempting to close the communication gap between deaf and hearing people.

3. Proposed Methodology

The method we used to evaluate the Google new Teachable machine involves employing a systematic approach consisting of the following steps (3.1 to 3.5) where the 3.1 to 3.2 is about how the data collection for ASL deployment done and the pre-processing that was taken place before training it with Google's teachable machine. 3.3 talks about the how the model was trained using Teachable Mahine,
3.4 describes the Teachable Machine and its capabilities and in the final section 3.5 we do the testing of the signs and further on evaluate the Teachable Machine's model.

3.1 Data Collection

For collecting data of each American Sign Language (ASL) a Python code with the help of OpenCV library was used for the video capturing and the 'HandDetector' Module from 'cvzone' for hand detection. Each frame that webcam is considered to identify the presence of hand in it, and once detected the bounding box around hand is calculated. To ensure we have uniformity in the dataset that we prepare the hand region is cropped with a padding offset and resized to predefined image size. Almost 300 images were generated and captured for each ASL letter's sign. These captured images are then saved to a specific folder with appropriate and unique timestamps, thereby making a labelled dataset. This also ensures the variety of hand sign postures and positions available for the training and evaluating of the Teachable machines

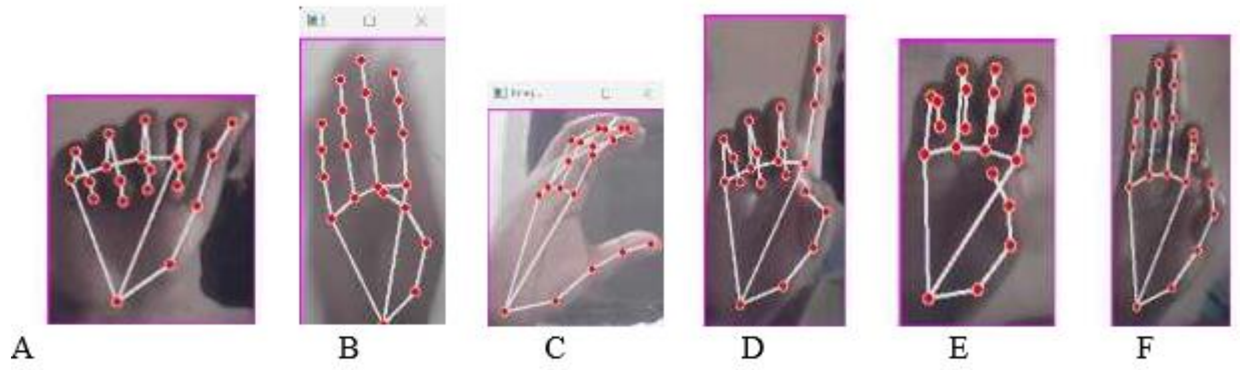

Fig. 1 The set of images are a sample of the captured images for our imagery dataset with labels indicating the which letter the sign represents

3.2 Data preprocessing

In this stage the captured images are made to undergo a step to enhance the quality and consistency of the captured imagery dataset of the American Sign Language. The cropped images containing the sign are resized and placed onto a background with a predefined dimension, this process makes that all the images have dimensions no matter of their original hand size or ratio. By making these changes we are simplifying the analysis and model training. The white also helps in isolating the actual gesture from the background noise, thereby making it to a standard such that the in the model training phase the model ability to identify the sign accurately is also benefitted. Once preprocessing is done the images become more capable of training a good model.

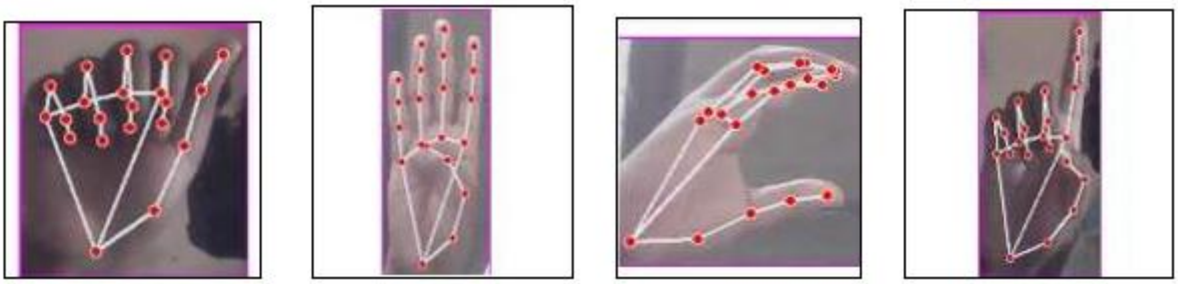

Fig. 2 The above images show a sample of the dataset which contains white padding in the background to bring all the images to a standard size.

3.3 Training using the Teachable machine

Here we involve all the 300 images made for each set of the sign. The imagery dataset was exported to the teachable machine platform. By leveraging the user-friendly interface of the tool. Once exported to the platform the Labelling of each of the data class of the images was done corresponding to ASL letter's sign it represented. Once the setup was prepared the Teachable machines built in capability were utilized to train the model and generate the model. During the training process, the Teachable Machine also provided Real-Time feedback on model performance allowing us to assess the condition while training.

3.4 About Teachable Machine and its capabilities

The Teachable Machine is a powerful and user-friendly web-based environment that helps broaden the access for playing field for machine learning training and modelling. It allows users to customize train models with their own dataset. It allows the user to export the data, label them and provides a seamless and easier way to produce ML models. Its capabilities go more than just making the model, it allows user to provide multiple types of datasets for training from images, audio and even sensor data. It removes the complexities of the machine learning algorithms and provides an easier and more convenient entry point. It also provides integration into web applications or other platforms, enabling seamless deployment and real-world applications, thereby making sure that that it acts as an 'Automatic ML Modeler'. It uses and leverages a pre-trained machine learning model and fine tunes them for specific tasks. This approach allows for faster training rates and better performance compared to going from starch al the way up to when dealing with limited data.

3.5 Testing the Model

Once the trained model is extracted from the teachable machine platform. A comprehensive testing phase has taken place. The testing code for the model was prepared using python and using a random selection of the ASL letters from the catalog of alphabet range from A-Z ensuring a diverse and unbiased test set. The user then displays the letter to the trained model through the webcam or by providing corresponding images. Now as the model processed the images and detected the letter, A box type frame occurs over the image with the letter that the model detects being displayed above the box. Meticulously we record the predictions, noting whether the predicted letter matched the actual input.

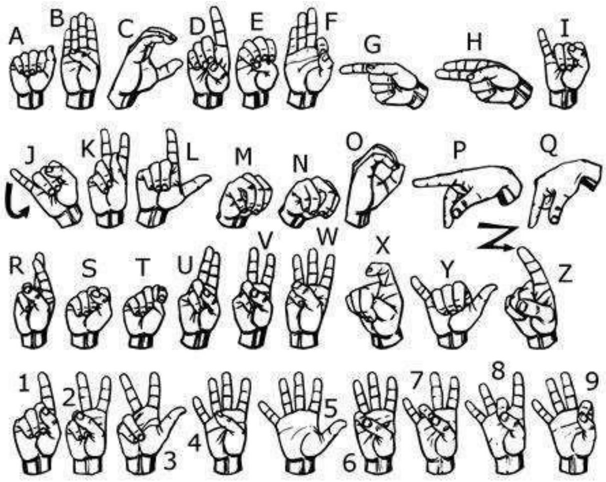

Fig. 3 The image represents the American Sign Language for 26 letters of the English alphabet

A Trail of 20 was selected for each letter to identify the model based on how long it takes to detect and whether the detected letter is actually the letter that is being displayed, as we want the model to detect the sign correctly and as quickly as possible which in turn also serves as a way to evaluate the model's accuracy or say effectiveness of the model or detection capacity.

Fig. 3 These images show how the model captures various letters when displayed to the webcam and displays the letter it detected.

4. Result Analysis and Discussion

Through the testing phase of the model, we have obtained a datasheet containing the data related to the model performance under the conditions we specified earlier. The table below represents the same

Table. 1 The table gives the recorded results of the testing phase the column named 'ASL Letters' displays each letter of the English alphabet column 2 represents the Correct Prediction out of 20 trails for each letter when the corresponding ASL sign was displayed and column 3 evaluates the Time taken for prediction for the corresponding letter

ASL Letter	Correct Prediction out of 20 trails	Time for Prediction
A	19	<1 sec
B	19	<1 sec
C	19	<1 sec
D	18	<1 sec
E	17	<1 sec
F	19	<1 sec
G	13	>2 sec
H	12	>2 sec
I	19	<1 sec
J	19	1-2 sec
K	18	<1 sec
L	18	<1 sec
M	12	>2 sec
N	13	1-2 sec
O	19	<1 sec
P	18	<1 sec
Q	18	<1 sec
R	17	>2 sec
S	14	>2 sec
T	13	1-2 sec
U	18	<1 sec
V	19	<1 sec
W	19	<1 sec
X	19	<1 sec
Y	19	<1 sec
Z	19	<1 sec

The results of the chi-square test provide valuable insights into the association between prediction time intervals and the accuracy of predictions

```
Chi-Square Statistic: 30.16481481481481
p-value: 0.0008050426421597845
Degrees of Freedom: 10
Expected Frequencies:
[[0.23076923 0.34615385 0.11538462 0.23076923 0.69230769 1.38461538]
 [1.38461538 2.07692308 0.69230769 1.38461538 4.15384615 8.30769231]
 [0.38461538 0.57692308 0.19230769 0.38461538 1.15384615 2.30769231]]
```

Fig.4 The image above shows the results of the Probabilistic-Test (Chi - Square) for evaluating the relation between the prediction time intervals and the accuracy

Chi-Square Statistic: The calculated chi-square statistic is approximately 30.165 from Fig. 4, indicates the strength of the association between prediction time intervals and accuracy. A higher chi-square value suggests a stronger association.
p-value: The p-value associated with the chi-square statistic is approximately 0.000805. This p-value is very small, indicating that the probability of observing the obtained results (or more extreme results) under the null hypothesis of no association is very low. In fact, it's less than the conventional significance level of 0.05. Therefore, we reject the null hypothesis and conclude that there is a statistically significant association between prediction time intervals and accuracy.
Overall, the results indicate that there is a significant association between prediction time intervals and the accuracy of predictions. This suggests that the prediction time intervals have an impact on the model's performance, and optimizing the prediction process could potentially improve accuracy. Further investigation into the underlying reasons for this association may be warranted to identify opportunities for improvement in the ML model.
Hence with these findings through the Chi-Square test we move on to now calculate the Average time for prediction and the Accuracy by prediction time thereby further dividing the total Correct Predictions by the Total trails we attain the accuracy of about 85.9%.

```
Results:
Accuracy: 0.8596153846153847

Average Time for Prediction:
1-2 sec: 3 trials
<1 sec: 18 trials
>2 sec: 5 trials

Accuracy by Prediction Time:
1-2 sec: 2.25
<1 sec: 16.7
>2 sec: 3.4
```

Fig. 5 The image represent the overall Accuracy Achieved and the Average Time for Prediction and the trails corresponding to the prediction tiume classes , also the Accuracy by Prediction Time value for each of the prediction time class.

Through the obtained analysis let us discuss merits and demerit about the Teachable Machine’s (Automatic ML modeler)
Some of the merits of the Platform is, it is one of a kind tool that provides easier to use interface and makes the machine learning and data model process much easier. The platforms capacity to handle the multiple variety of data input is a star feature, apart from this through the Result analysis and chi- square test the model performs well for small data class with around 5-6 and most of the predictions come out really good as for any automatic ML modeler to come up with a model with good detection without any human intervention to this level is really an appreciable aspect of the Teachable Machine tool. The tool as suggested by Google is in the initial aspects in the field of automatic modelling and is having a good number of improvements.
Some of the demerits of the models prepared by the platform through its feature of in-built automatic modeler is that though it performs well in detecting the classes of small dataset, but once size of the

dataset increases the model struggles to recognize the complex symbols and starts to mis-interpret the symbols. Similarly in our case, with the huge size of the dataset it was found that the model finds itself in a difficult place to identify the sign. Apart from the complexity of the sign the Teachable Machine still needs improvement to automatically model detection models as it fails to identify intricate points to in the sign so that it can detect the sign.

5. Conclusion

To conclude the Analysis of the Teachable machine with the help of assessing American sign language detection it can be said that the models produced by Teachable Machine are preforming with a accuracy of about 85.9% and for most of the requirements such as developing any small models using Machine learning it can prove to useful but cannot be guaranteed as such as the best as the model does face difficulty in recognizing the letter with similar kind of sign posture. The model also starts to struggle when the data volume increases, and the computational need goes up. Though it can be used for small model training for large model training it is still not suitable, and the technology needs to be improved. Another way that can be done to meet requirements is by training multiple small models with Teachable Mahine and combining it later to get a good improvement in the detection is an alternate approach.

References

1. Bhatt, R., Malik, K. and Indra, G., 2024, March. ASL Detection in Real-Time using TensorFlow. In 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI) (Vol. 2, pp. 1-6). IEEE.
2. Mittal, A., Kumar, P., Roy, P.P., Balasubramanian, R. and Chaudhuri, B.B., 2019. A modified LSTM model for continuous sign language recognition using leap motion. IEEE Sensors Journal, 19(16), pp.7056-7063.
3. Rahman, M.M., Kurtoglu, E., Mdrafi, R., Gurbuz, A.C., Malaia, E., Crawford, C., Griffin, D. and Gurbuz, S.Z., 2021, June. Word-level ASL recognition and trigger sign detection with RF sensors. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8233-8237). IEEE.
4. Gurbuz, S.Z., Gurbuz, A.C., Malaia, E.A., Griffin, D.J., Crawford, C.S., Rahman, M.M., Kurtoglu, E., Aksu, R., Macks, T. and Mdrafi, R., 2020. American sign language recognition using rf sensing. IEEE Sensors Journal, 21(3), pp.3763-3775.
5. Shi, B., Brentari, D., Shakhnarovich, G. and Livescu, K., 2021. Fingerspelling detection in american sign language. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4166- 4175).
6. Deep, A., Litoriya, A., Ingole, A., Asare, V., Bhole, S.M. and Pathak, S., 2022, August. Realtime sign language detection and recognition. In 2022 2nd Asian Conference on Innovation in Technology (ASIANCON) (pp. 1-4). IEEE.
7. Joshi, A., Sierra, H. and Arzuaga, E., 2017, August. American sign language translation using edge detection and cross correlation. In 2017 IEEE Colombian Conference on Communications and Computing (COLCOM) (pp. 1- 6). IEEE.
8. Jiang, X. and Ahmad, W., 2019, August. Hand gesture detection based real-time american sign language letters recognition using support vector machine. In 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on

Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech) (pp. 380-385). IEEE.
9. Sharma, Shikhar, and Krishan Kumar. "ASL-3DCNN: American sign language recognition technique using 3-D convolutional neural networks." Multimedia Tools and Applications 80, no. 17 (2021): 26319-26331.
10. Abdulhussein, A.A. and Raheem, F.A., 2020. Hand gesture recognition of static letters American sign language (ASL) using deep learning. Engineering and Technology Journal, 38(6A), pp.926-937.
11. Lee, C.K., Ng, K.K., Chen, C.H., Lau, H.C., Chung, S.Y. and Tsoi, T., 2021. American sign language recognition and training method with recurrent neural network. Expert Systems with Applications, 167, p.114403.
12. Abiyev, R.H., Arslan, M. and Idoko, J.B., 2020. Sign language translation using deep convolutional neural networks. KSII Transactions on Internet and Information Systems (TIIS), 14(2), pp.631-653.
13. Joze, H.R.V. and Koller, O., 2018. Ms-asl: A large-scale data set and benchmark for understanding american sign language. arXiv preprint arXiv:1812.01053.
14. Shi, B., Del Rio, A.M., Keane, J., Michaux, J., Brentari, D., Shakhnarovich, G. and Livescu, K., 2018, December. American sign language fingerspelling recognition in the wild. In 2018 IEEE Spoken Language Technology Workshop (SLT) (pp. 145-152). IEEE.
15. Sanmitra, P.R., Sowmya, V.S. and Lalithanjana, K., 2021. Machine Learning Based Real Time Sign Language Detection. International Journal of Research in Engineering, Science and Management, 4(6), pp.137-141.
16. Fang, B., Co, J. and Zhang, M., 2017, November. Deepasl: Enabling ubiquitous and non-intrusive word and sentence-level sign language translation. In Proceedings of the 15th ACM conference on embedded network sensor systems (pp. 1-13).