ISSN 2394-739X

IJRSET JUNE Volume 10 Issue 6

International Journal for Research in Science Engineering & Technology (IJRSET)

https://www.doi.org/10.5281/zenodo.8382974

Sign Language Recognition Utilizing LSTM & Media pipe for Dynamic Gestures of ISL

[1] P. Swetha, [2] K. Sucharitha
[1, 2] Assistant Professor,
[1, 2] Department of Computer Science Engineering,
[1, 2] St. Martins Engineering College, Secunderabad, Telangana.

ABSTRACT: Humans, in general, are social creatures who communicate themselves through an assortment of spoken languages. Deaf and Mute individuals converse in a manner that's comparable, however many others areignorantoftheirsignlanguage. Asaresult, thereisaneedtodevelopasystemthatfacilitatescommunication among the hearing and hard-of-hearing communities. This research offers a real-time Indian Sign Language (ISL) recognition system for 24 dynamic signals using the Mediapipe framework and an LSTM network. The method proposed in the study involves training a LSTM to differentiate between different signs using a dataset created of 24 dynamic gesture signs.To accomplish dataset creation, a pre-trained Holistic model of the Mediapipe framework is used as a feature extractor. Theresultsof thestudy demonstrate that theaboveapproachachieves 97%testaccuracy.

Keywords: [Indian Sign Language, Dynamic Gestures, Mediapipe, LSTM, Computer Vision.]

1. INTRODUCTION

One of the most crucial pillars of daily life is communication because it allows people to express theirideas and opinions and thus helps them integrate into society. The ability to hear and speak, however, isnot shared by all people, and thereby some find it difficult to use. As a result, they are unable tocommunicatenormally and struggleto fit into society.

Sign language recognition (SLR) is crucial in the field of assistive technology for persons with hearingimpairments. This technology enables seamless communication and access to diverse services for thisdemographic.Peoplewithhearingimpairmentsmayhavesubstantialdifficultiesintheirdailyactivitiesif they do not have proficient sign language recognition technology. They may struggle to communicate the irneeds, understandinformation, orevenparticipateinsocialactivities. Bydevelopingandimprovingsignlanguagerecognitiontechnology, wecanempowertheseindividualstoleadmoreindependent and fulfilling lives, bridging the communication gap and promoting inclusivity within society.

The goal is to study the usage of LSTM (Long Short Term Memory) networks in the recognition ofIndian Sign Language (ISL) at the word level. The system's primary job is to quickly and correctly detect, categorise and translate the signs performed in ISL by utilizing neural networks and Computer Vision. The present system in this study can currently handle the recognition of up to 24 ISL vocabularywordsin real-time withinamatterofafew seconds.

Gesture-based sign language recognition systems face numerous hurdles, particularly in the context of Indian Sign Language (ISL) where the alphabets are widely different from American Sign Language (ASL) asshownin Figure1.Theprimaryobstaclestemsfromtheintricateanddynamichandmovements integral to ISL. Another issue pertains to the diversity in signing approaches among different individuals, whichcomplicatesthedevelopmentofauniversallyapplicablerecognitionmodel.Furthermore, the existence of background disturbances and obstructions amplifies the challenge ofachievingpreciserecognition.Nevertheless,theadoptionofLongShort-TermMemory(LSTM)networkshasdemonstratedpromisingoutcomesinenhancingtheaccuracyandefficiencyofISLrecognition systems.

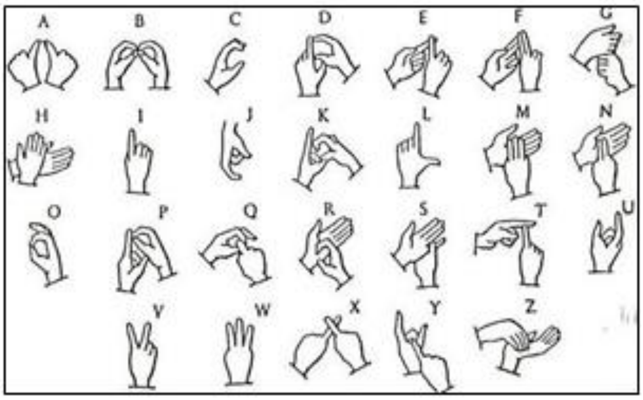

Figure1: Indian Sign Language Alphabets

2. LITERATURESURVEY

Dhivyasri S, et al paper [1] proposes the use of SURF (Speeded Up Robust Feature) method for feature extractions. For identifying and describing local features in images, SURF is a well-liked computer vision technique. The technology can track the movements of important sign language gestures in real-time by using SURF to recognize their key points. As a result, the technology will be better able todistinguishbetween similar signs and recognisesign language more accurately.

Furthermore, the examination of existing literature reveals multiple endeavours in the advancement of systems for recognizing sign language. These efforts encompass a range of approaches, such as CNN,RNN, SVM, and K-Means for SLR, in addition to the application of SVM and CNN for translating textintogestures [2].

The majority of constraints inherent in the study carried out in [1] [2] pertains to the utilization of static and isolated gestures. To advance this field, further investigation is essential to incorporate the subject of

ISSN 2394-739X
IJRSET JUNE Volume 10 Issue 6
dynamicgesturerecognition, encompassingtheidentificationofmotionaswellasshiftsinhandconfiguration and orientation across temporal dimensions. This aspect is significant because dynamicmovements within sign language play an essential role in aiding the translation of subtle interpretations.Furthermore, improvementstothetechnology'saccuracyandreliabilityarerequiredtoensureitsusefulnessin real-worldcircumstances.

Purva Chaitanya Badhe, et al paper [3] employs a hand-crafted feature extraction technique. Theyintroduced a procedure for recognizing Indian Sign Language (ISL) using a vision-oriented approach.Themethodputforthdeviatesfromexistingapproaches byemployinganRGBimagecenteredstrategy, as opposed to alternatives which rely on depth images or data from a leap motion sensor. They use anartificialneuralnetworkfortheclassificationofthegestures.Thetrainingaccuracyisaround98%.Sinceit is a small dataset, thevalidation accuracyis 63%.

To enhance the precision as well as dependability of the technology, it might be imperative to augmentthe scale of the dataset utilized for both training and validation purposes. This extension would facilitatetheincorporationofamoreextensivearrayofhandgestures,therebyguaranteeingthesystem'sadeptness in accurately discerning even nuanced fluctuations in manual movements. Furthermore, theadoption of more sophisticated machine learning approaches, which include deep neural networks [4],has the potential to enhance the system's accuracy. However, the use of these strategies may needconsiderablecomputingresourcesaswellasacomprehensiveunderstandingofimplementationtechniques.

Deep R. Kothadiya, et al used a vision transformer to recognise static Indian signs in their paper [5]. Theproposed method divides the sign into a series of positional embedding patches, that are subsequentlyprocessed by a transformer block with four self-attention layers and amultilayer perceptron network.Theempiricalfindingsdemonstratethatavarietyofaugmentationtechniquesyieldsatisfactoryrecognition of gestures. Moreover, the method put forth in this study requires only a relatively limitedquantityof training epochs to achieveanaccuracylevel of 99.29%.

TheuseofvisiontransformersintherecognitionofstaticIndiansignlanguageisapromisingdevelopment in the field of gesture recognition based on research carried out in [6]. By de-constructingthe indicators into positional embedding patches and leveraging a transformer block equipped with self-attentionmechanisms,themodeldemonstratesexceptionalprecisionwithlimitedtrainingdata.Nevertheless, the task persists in terms of adapting this approach to discern dynamic gestures andseamlessly integrating it into operational systems that work in real-time. Additional investigation iswarranted toprobe the latentcapacities of thesesophisticated methodsin the enhancement of systemsforgesturerecognition.

Muhammad Al-Qurishi, et al propose a general framework for researchers in their paper [7], whichdiscusses their relative strengths and weaknesses.This investigation also demonstrates the significanceof input modalities within this domain. Evidently, the utilization of diverse data sources encompassingvisual-oriented as well as sensor-oriented channels exhibits superior performance compared to a uni-modal analysis.. Furthermore, recent advancements have enabled researchers to progress from simplerecognition of sign language characters and words to the ability to translate continuous sign languagedialogue with minimal delay. Many of the models mentioned are relatively effective for a variety oftasks, but none currently have the generalisation potential required for commercial deployment. Onemajor complication found in the study pertains to the matter of individual divergence in gestures, whichcan lead to inconsistencies in recognition accuracy. To address this, machine learning algorithms can beproposed to adapt to individual users' unique gestures over time [8]. Another challenge is the need forrobustnessinreal-worldsettings,wherelightingconditionsandbackgroundcluttercanaffectrecognition performance. To surmount this challenge, certain scholarly inquiries have delved into theutilization of depth sensors and three-dimensional cameras to amass more intricate insights concerninggestures.

Maher Jebali, et al describe a computer vision-based system for recognising signs in a continuous signlanguage clip in their paper [9].The system isdivided into two stages: sign wordextraction andcategorization. Isolating sign words from video frames is the most difficult task in this process. Theyoffer an innovative algorithm capable of detecting appropriate word boundaries in a continuous signlanguage video for this goal. This algorithm is used to extract isolated signs from video, utilizing bothhandstructure and motion characteristics. It show casesenhancedperformanceincontrasttootherpreviously published endeavours in the same domain. The extracted signs are categorized and identified using the Hidden Markov Model (HMM) in the recognition stage, which was strongly embraced afterassessing HMM with other methodologies like Independent Bayesian Classifier Combination (IBCC)[10].Thesystemfunctionsadmirably,exhibitingarateofrecognitionof95.18%forone-handedmotions and 93.87% for two-handed gestures.When using head poseand eye gaze attributes, theframework attains 2.24% and 2.9% improvement on one and two hand gestures, respectively, whencompared to systems just using manual attributes. These findings are based on a dataset of 33 isolatedsigns.

Ilias Papastratis, et al in their paper [11] offer an innovative framework that leverages the syntacticalstructure of oral communication. This novel approach is constructed upon the linguistic patterns gleanedfrom a sizeable corpus of text sentences. The framework is comprised of three primary modules: cross-modal re-ranking, conditional sentence generation, and word existence verification. By conducting aseries of parallel binary classifications to check the occurrence of the terms in the lexicon, they then put the terms together and used a pre-trained speech generator to generate candidate sentences in the spokenlanguagevariation. Usingacross-modalre-rankingmodel, the translationoutcomethatismostsemantically identical to the original sign video is chosen. The assessment of the framework is doneusing the CSL and RWTHPHOENIX-Weather 2014 T which are SLT benchmarks [12]. Experimentalfindingsdemonstrated thatthesuggested frameworkperformed commendablyon both datasets.

E Rajalakshmi, et al [13] created a novel, natural, multi-signer Indo-Russian Sign Language databasecomprising isolated sign gestures. They use a multi-semantic discriminative feature learning deep neuralnetwork [14] and spatial, temporal and sequential feature learning method for SLR. The limitation is thatthenewly created datasetis constrained to staticisolated sign languagegestures.

In their academic research, Z. Wang et al have put forth a unique proposal. They have introduced aningeniousframework,characterizedbyanattention-centricencoder-decodermodel,synergisticallypaired with a multi-channel convolutional neural network (CNN) [15]. Notably, this methodology hingesupon the strategic deployment of wearable armbands, thoughtfully embedded with an array of sensors.These specialized devices are meticulously fastened onto the forearms, strategically poised to adeptlyapprehend a dual spectrum of actions: encompassing both sweeping arm movements and the intricatenuances of finger motions. The dataset is quite small and this methodology requires wearable devices,sensorssuch as leap-motion and kineticdevices [16].

In the realm of sign language recognition, the work carried out in paper [17][18] describes an innovative approach which leverages robust deep learning techniques to tackle the intricate task of sign languageinterpretation. Thisapproachinvolvestheutilizationoftexturemapstointricately encodebothhand location and motion aspects. Impressively, their devised model achieved a commendable accuracy levelof 87.02%. However, a notable challenge that surfaced within this model pertained to its recognitionaccuracy when dealing with signs that exhibit similarity in form. Despite its considerable success, themodelencountered difficultiesin accuratelydistinguishing betweensuch closelyrelated signs.

The approach introduced by Neel Vasani, et al [19] centers around the transformation of sentences intoconcise notations (referred to as gloss). These notations are then employed to generate synthetic videoframes by the method discussed in the word carried out in [20], using a Generative Adversarial Network(GAN) architecture. This process culminates in the creation of a visual representation for the given inputsentence. This intricate interplay of techniques ultimately leads to the development of a comprehensivevideo depiction corresponding to the original input sentence. The issue with respect to this study is thatdatasetofvideos isnotinhigh-resolution,andtraining requiresa many epochsandcomputationalpower.

P.V. V Kishore, et al used Artificial Neural Networks to categorize and detect signers' movements fromvideo frames [21]. The rate of recognition of around 93% was achieved in their approach. The drawbackwas the usage of limited dataset with low resolution for faster computing. Furthermore, the approachdoesnot takeinto account continuoussign languageidentification in real time.

Anjan Kumar Talukda, et al proposed a vision-based continuous Sign Language spotting system, buildaround a two-state Hidden Markov model (HMM) with Gaussian emission probability [22]. They wereable to achieve an accuracy of around 83%. Exclusively the dataset containing videos of American SignLanguage is employed as the primary resource within this research. Additionally, it was only capable ofspottingonesign in a video.

Pan Xie, et al in their paper [23] propose a novel content-aware neighbourhood gathering method toselect relevant features dynamically and disentangled relative position encoding (DRPE) method for therelative position information to SLTR model. The scope of their study was constrained to encompasssolelythedomains of German and ChineseSign Languages [24].

In their research endeavour, Jian Zhao, et al [25] introduced an innovative framework founded upon thevalidation of word presence, subsequent sentence generation, and a process of cross-modal re-ranking inthe realm of Signed Language Translation (SLT). The scope of their study was predominantly confinedto the domain of Chinese Sign Language and specifically focused on the RWTH-PHOENIX-Weather2014 T dataset. However, a notable limitation that came to the forefront in theirwork was associatedwithwordpresencevalidationmechanism,whichexhibitedshortcomingsinaccuratelyidentifyingpivotalcontent words.

3. CONCEPTSANDTHEORIESBEHINDSLR

3.1 Computer Vision

Computer visionhas the capacity to capture and process visual data from video streams, specifically thehand gestures made by signers. The methodologies are initially employed to identify and capture handgestures, whicharesubsequentlysubjectedtoanalysisfortheextractionofpertinentcharacteristics

ssentialfortheclassificationprocess.Thesefeaturescanincludehandshape, orientation, and movementpatterns.

3.2 Recurrent Neural Network

RNNs have a feedback mechanism that allows information to be passed from one step of the sequence tothe next as shown in Figure 2. The vanishing gradient problem is a significant disadvantage of classicalRNNs, where the gradients used to update the network's weights become very small, making it difficultforthe network to learn long-term dependencies.

To address this issue, many RNN variations, such as Long Short-Term Memory (LSTM) and GatedRecurrent Units (GRU), have been created, which contain additional methods to limit the propagation ofinformationthrough the networkand preventthevanishing gradient problem [26].

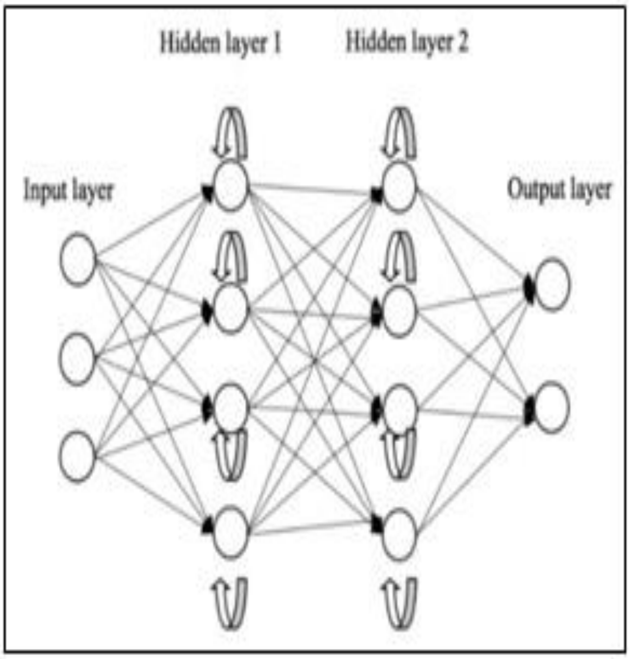

Figure2: RecurrentNeuralNetwork

3.3. LongShortTermMemoryNetwork

LSTMshaveanelaboratedesignthatincorporates"memorycells" and"gates"thatgovernthemovement of data across the network, as opposed to traditional RNNs, that utilize abasic feedback looptosend information gained from one time step to subsequent one.

An LSTM receives an input vector as well as a hidden state vector containing data from the precedingtimestepateachtimestep.Theinputandhiddenstatevectorsarethenprocessedbythenetworkthrough a set of gates that

govern the information that flows through and out of the memory cells. Thegates are composed of sigmoid functions which return values that span 0 to 1, indicating which elementsoftheinputinformation aswell as hiddenstate vectorsmust be permittedinto thememory cells.
Memory cells store data over multiple time steps, enabling the network to detect long-term dependenciesin input data. The LSTM output is a combination of the current memory cell state as well as the hiddenstatevectorat eachtimestep, andit has thepotential to beused forpredictionorclassification.

4. IMPLEMENTATION

The implementation work- flow for Dynamic Sign Gesturerecognitionis as shown in theFigure3.

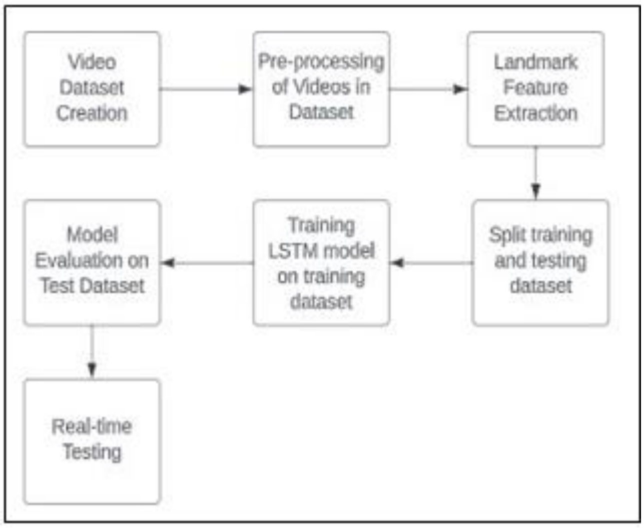

Figure 3: Block Diagram of Dynamic ISL Gesture Recognition

4.1 Dataset Creation and Pre-Processing

Each dynamic sign language gesture was captured on film 30 times for the dataset utilized in thisinvestigation. Each video is made up of 30 frames which was necessary in order completely film the signgesture.Insignswhichdidn’tneedall30frames,they were augmented toinclude zeros sothatitreached30 frames.

4.2 Feature Extraction

For capturing the required features that make up the dynamic sign gesture, certain landmarks were takeninto consideration. The Mediapipe framework was employed to pick the right and left hand, face, andpose markers for this operation. In total there are 543 feature landmarks extracted which has 33 poselandmarks, 468 face landmarks, and 21 hand landmarks for right and left hand each. Since dynamicgestures include more than just hand movement, the Holistic model was utilized which includes all threemodelsto determinetheco-ordinates of landmarkson hand, poseand face.
Thesefeaturesarestoredinanumpyarrayfileforeachframeofeach video.Sinceitisthree-dimensional, the x, y and z co-ordinates are considered. Thereby, each numpy array file has a total of1662features.

4.3 Training the Model

Split the preprocessed dataset into training and validation sets based on Table I values. Train the modelwhichis depicted in Figure4, on thetraining set usingtheextractedfeatures.

```
Layer (type)              Output Shape           Param #
=================================================================
lstm_9 (LSTM)             (None, 30, 64)         442112

lstm_10 (LSTM)            (None, 30, 128)        98816

lstm_11 (LSTM)            (None, 64)             49408

dense_9 (Dense)           (None, 64)             4160

dense_10 (Dense)          (None, 32)             2080

dense_11 (Dense)          (None, 24)             792

=================================================================
Total params: 597368 (2.28 MB)
Trainable params: 597368 (2.28 MB)
Non-trainable params: 0 (0.00 Byte)
```

Figure4: Model Summary

Basedonthevaluesin Table I whichdescribesthehyper-parametersdata, the modelwasfine tunedtoenhanceperformance.

Hyper-Parameters	Values
TrainingData	80%(576videos)
TestingData	20%(144videos)
SequenceLength	30
LSTMLayers+Neurons PerLayers	3Layers 64neurons 128neurons 64neurons
DenseLayers+ Neurons perlayers	3Layers 64neurons 32neurons 24neurons
ActivationFunction -inputandhiddenlayers	Relu
ActivationFunction -Outputlayer	Softmax
Optimizer	Adam
BatchSize	128
Epoch	250

Table1: ModelHyper-Parameters

4.4 ModelEvaluation

Assessthemodel'sperformancewithevaluationmetrics,usingthe validationsettodetermineitsaccuracy, precision, recall, and F1-score. The hyper-parameters and network architecture were adjustedtooptimizeperformance

4.5 Real-TimeTesting

The trained model was later deployed on real-time scenarios to evaluate its performance recognisingDynamicSign Gestures.

5. Results And Analysis

The SLR system for dynamic gestures was able to achieve a training accuracy of 98.5% within 250epochs. The model was trained employing different batch sizes, starting from the default 32, 64 and 128.Themodel with batch sizeof 128 performedbetterin comparison to theothers.
After 250 epochs the model’s accuracy started dropping with increase in loss. This indicates that themodel reached its optimal performance after 250 epochs and any further training did not yield significantimprovements.The128batch sizeproved tobe effectiveinachievinghigh accuracyand efficiency.
The modelachievedanaccuracy of 97%

ISSN 2394-739X
IJRSET JUNE Volume 10 Issue 6

afterbeingtestedonthevalidationdataset.
The confusion matrix of each dynamic gesture is as depicted in Figure 5. The precision, recall and f1-scoreofeach dynamicsign gestureis calculated as shown asFigure7.

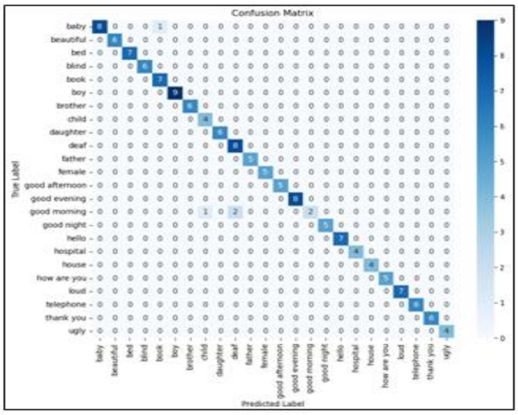

Figure5: ConfusionMatrix

The average precision, recall and F1-score for validation set was 98%, 97%and 97% as depicted inFigure6.

	precision	recall	f1-score	support
baby	1.00	0.89	0.94	9
beautiful	1.00	1.00	1.00	6
bed	1.00	1.00	1.00	7
blind	1.00	1.00	1.00	6
book	0.88	1.00	0.93	7
boy	1.00	1.00	1.00	9
brother	1.00	1.00	1.00	6
child	0.80	1.00	0.89	4
daughter	1.00	1.00	1.00	6
deaf	0.80	1.00	0.89	8
father	1.00	1.00	1.00	5
female	1.00	1.00	1.00	5
good afternoon	1.00	1.00	1.00	5
good evening	1.00	1.00	1.00	8
good morning	1.00	0.40	0.57	5
good night	1.00	1.00	1.00	5
hello	1.00	1.00	1.00	7
hospital	1.00	1.00	1.00	4
house	1.00	1.00	1.00	4
how are you	1.00	1.00	1.00	5
loud	1.00	1.00	1.00	7
telephone	1.00	1.00	1.00	6
thank you	1.00	1.00	1.00	6
ugly	1.00	1.00	1.00	4
accuracy			0.97	144
macro avg	0.98	0.97	0.97	144
weighted avg	0.98	0.97	0.97	144

Figure6: Precision, Recall, F1-Score, Support of Test Dataset

These findings show the possibility of using SLR systems for dynamic gestures in ISL identification, pavingtheway toward further study and advancement in this field.
Furthermore the modelwasalsotestedwithreal-time signgesturesusinga standard laptop camera where the gestures were done with no sign gesture isolation. The figure 7 and 8 are some examples ofdynamicsignsbeingrecognizedbytheSLRsystem.Themodel wasabletorecognizethesignsaccurately but when switching from one to next, the response led to few false positives since the camera was picking up each change and trying to recognize the gesture. This suggests that exists a future scopetocarry out further research with respect to successivedynamicsign recognition.

Figure7: Testing in real time for word -Beautiful

Figure8: Testing in real time for word-Ugly

Conclusion and Future Work

In conclusion, the study successfully developed and evaluated a SLR system for dynamic gestures in ISLrecognition. The results indicate that the model's performance reached its peak after 250 epochs with abatchsizeof128.Furtherresearchcanfocusonexploringdifferentarchitecturesandhyper-parameterstopotentially improveaccuracyevenfurther.
Additionally, investigating the use of SLR systems for other sign languages and expanding the datasetcould yield valuable insights and advancements in the field of gesture recognition. Furthermore, it wouldbe interesting to investigate the impact of incorporating temporal information into the SLR system fordynamic gestures.Thiscouldinvolve exploringrecurrentneuralnetwork architecturesorattentionmechanismsto capturethesequential natureof sign language.
Moreover, conducting user studies to evaluate the usability and effectiveness of the SLR system in real-worldscenarios would providevaluablefeedbackforimproving its practicalapplications.
Overall,thefindings fromthisstudylaythefoundationforfurtherstudiesinthedomainofISLrecognitionandpavepathwaysforthecreationofmorerobust and accurategesturerecognitionsystems.

REFERENCES

[1]. D. S, K. H. K B, A. M, S. M, D. S and K. V, "An Efficient Approach for Interpretation of IndianSign Language using Machine Learning," 2021 3rd International Conference on Signal ProcessingandCommunication(ICPSC),Coimbatore,India,2021,pp.130-133,doi:10.1109/ICSPC51351.2021.9451692.
[2]. K.Shenoy,T.Dastane,V.RaoandD.Vyavaharkar,"Real-timeIndianSignLanguage(ISL)Recognition," 2018 9th International Conference on Computing, Communication and Networking Technologies(ICCCNT), Bengaluru, India,2018, pp.1-9,doi:10.1109/ICCCNT.2018.8493808.
[3]. P. C. Badhe and V. Kulkarni, "Artificial Neural Network based Indian Sign Language Recognitionusinghandcraftedfeatures,"202011thInternational ConferenceonComputing,CommunicationandNetworkingTec

hnologies(ICCCNT),Kharagpur,India,2020,pp.1-6,doi:10.1109/ICCCNT49239.2020.9225294.

[4]. A. Wadhwan, and P. Kumar, “Sign Language Recognition Systems: A Decade Systematic LiteratureReview”,ArchivesofComputationalMethodsinEngineering,Springer,2019,DOI:https://doi.org/10.1007/s11831-019-09384-2

[5]. D. R. Kothadiya, C. M. Bhatt, T. Saba, A. Rehman and S. A. Bahaj, "SIGNFORMER: DeepVisionTransformer for Sign Language Recognition," in IEEE Access, vol. 11, pp. 4730-4739, 2023, doi:10.1109/ACCESS.2022.3231130.

[6]. K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y.Zhang, and D. Tao, ‘‘A survey on vision transformer,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol.45,no. 1, pp. 87–110, Jan. 2023.

[7]. M. Al-Qurishi, T. Khalid and R. Souissi, "Deep Learning for Sign Language Recognition: CurrentTechniques, Benchmarks, and Open Issues," in IEEE Access, vol. 9, pp. 126917-126951, 2021, doi:10.1109/ACCESS.2021.3110912.

[8]. S.Stoll, N.C.Camgoz, S.Hadfield, and R.Bowden,‘‘ Text 2sign:Towardssignlanguageproduction using neural machine translation and generative adversarial networks,’’ Int. J. Comput.Vis., vol. 128, no. 4, pp. 1–18, 2020.

[9]. Jebali, M., Dakhli, A.& Jemni, M. “Vision-basedcontinuous sign language recognition usingmultimodalsensorfusion,”EvolvingSystems,vol12,1031–1044(2021).https://doi.org/10.1007/s12530-020-09365-y

[10]. Wenwen Yang, Jinxu Tao, Zhongfu Ye, “Continuous sign language recognition using level buildingbased on fast hidden Markov model, Pattern Recognition Letters, vol 78, pp 28–35, ISSN 0167-8655,https://doi.org/10.1016/j.patrec.2016.03.030.

[11]. I. Papastratis, K.Dimitropoulos, D.Konstantinidis and P.Daras," Continuous Sign Language Recognition Through Cross-Modal Alignment of Video and Text Embeddings in a Joint-LatentSpace,"inIEEEAccess,vol. 8,pp. 91170-91180,2020, doi:10.1109/ACCESS.2020.2993650.

[12]. O. Koller, J.Forster, and H.Ney,‘‘ Continuous sign languagerecognition:Towardslargevocabularystatisticalrecognitionsystemshandlingmultiplesigners,’’Comput.Vis.ImageUnderstand.,vol. 141, pp. 108–125, Dec. 2015.

[13]. E.Rajalakshmietal.," Multi-Semantic Discriminative Feature LearningforSignGestureRecognition Using Hybrid Deep Neural Architecture," in IEEE Access, vol. 11, pp. 2226-2238,2023,doi: 10.1109/ACCESS.2022.3233671.

[14]. X. Jiang, M. Lu, and S.-H. Wang, ‘‘An eight-layer convolutional neural network with stochasticpooling, batch normalization and dropout for fingerspelling recognition of Chinese sign language,’’MultimediaTools Appl.,vol.79, nos. 21–22, pp. 15697–15715, Jun. 2020.

[15]. Z. Wangetal.,"HearSignLanguage:AReal-TimeEnd-to-EndSignLanguageRecognitionSystem," in IEEE Transactions on Mobile Computing, vol. 21, no. 7, pp. 2398-2410, 1 July 2022,doi:10.1109/TMC.2020.3038303.

[16]. G. Marin, F. Dominio, and P. Zanuttigh, “Hand gesture recognition with leap motion and kinectdevices,”in Proc.IEEEInt.Conf. ImageProcess.,2014,pp. 1565–1569.

[17]. E. Esco be do, L.Ramirez and G. Camara," Dynamic SignLanguageRecognitionBasedonConvolutionalNeuralNetworksandTextureMaps,"201932ndSIBGRAPIConferenceonGraphics,PatternsandImages(SIBGRAPI),RiodeJaneiro,Brazil,2019,pp.265-272,doi:10.1109/SIBGRAPI.2019.00043.

[18]. H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould, “Dynamic image networks for actionrecognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016, pp.3034–3042.

[19]. N. Vasani, P. Autee, S. Kalyani and R. Karani, "Generation of Indian sign language by sentenceprocessing and generative adversarial networks," 2020 3rd International Conference on IntelligentSustainableSystems(ICISS),Thoothukudi,India,2020,pp.1250-1255,doi:10.1109/ICISS49785.2020.9315979.

[20]. Stoll, S., Camgöz, N. C., Hadfield, S., & Bowden, R. (2018, September). “Sign language productionusing neural machine translation and generative adversarial networks.” In Proceedings of the 29thBritishMachine Vision Conference(BMVC 2018). BritishMachine Vision Association.

[21]. P. V. V. Kishore, A. S. C. S. Sastry and A. Kartheek, "Visual-verbal machine interpreter for signlanguage recognition under versatile video backgrounds," 2014 First International Conference onNetworks&SoftComputing(ICNSC2014),Guntur,India,2014,pp.135-140,doi:10.1109/CNSC.2014.6906696.

[22]. A.K.Talukdar and M.K.Bhuyan," Vision-Based Continuous Sign LanguageSpottingUsingGaussian Hidden Markov Model," in IEEE Sensors Letters, vol. 6, no. 7, pp. 1-4, July 2022, Art no.6002304,doi: 10.1109/LSENS.2022.3185181.

[23]. P.Xie, M.Zhaoand X.Hu," PiSLTRc: Position-Informed Sign Language Transformer With Content Aware Convolution," in IEEE Transactions on Multimedia, vol. 24, pp. 3908-3919, 2022,doi:10.1109/TMM.2021.3109665.

[24]. J. Zhang, W. Zhou,X. Chao, J. Pu, and H. Li, “Chinese sign language recognition with adaptiveHMM,”in Proc.IEEEInt.Conf. MultimediaExpo., 2016, pp.1–6.

[25]. J. Zhao, W. Qi, W. Zhou, N. Duan, M. Zhou and H. Li, "Conditional Sentence Generation andCross-Modal Reranking for Sign Language Translation," in IEEE Transactions on Multimedia, vol.24,pp. 2662-2672, 2022, doi: 10.1109/TMM.2021.3087006.

[26]. F. Obaid, A. Babadi, and A. Yoosofan, “Hand Gesture Recognition in Video Sequences Using DeepConvolutionalandRecurrentNeuralNetworks,”Appl.Comput.Syst.,vol.25,no.1,pp.57–61,May2020, doi: 10.2478/acss-2020-0007.