SIGNLANGUAGETRANSLATIONBASEDONHANDGESTURESCONVERTINGINTOTEXTAND SPEECHIN ENGLISHAS WELLAS TAMILUSINGRANDOMFORESTALGORITHM

GayathriM[#1], KaviyaK[#2], PriyankaS[#3],Dr.MadhurikkhaS[#4]
[#1,2,3]Department Of ComputerScienceAnd Engineering,
[#4]Associate Professor / AI&DS
JayaSakthi EngineeringCollege,Chennai,602024,TamilNadu,India
gayathri30kavi@gmail.com[#1], kaviyak98240@gmail,com[#2], js.priyanka06@gmail.com[#3],
madhurikkha@gmail.com[#4]

Abstract— A real-time system that can decipher sign languagefrom a live webcam stream is presented by the Sign LanguageConversion Project. Using the powerful landmark identificationfeatures of the Mediapipe library, the project takes importantdata out of every frame, including hand landmarks. Followingtheirdetection,theselandmarkcoordinatesarecollatedan dsaved in a CSV file for later examination. This landmark data isused to train a Random Forest Classifier, which uses machinelearning techniques to identify different sign language patterns.Thetrainedmodelpredictsthesignlanguageclassanditspro bability in real-time as it processes the camera data. In ordertoimproveaccessibility,thesystemoffersEnglishtranslationide as for the identified signs, which are accompanied by voicetranslations into Tamil and English. The video feed has thesetranslations superimposed on it, providingreal-time translationsand visual cuesenhancesaccessibilityfor individuals who aredeaf or hard of hearing, allowing them to communicate moreeffectivelywithotherswhomay notunderstandsignlanguage.

Keywords— **Random forest algorithm, Mediapipe,sign languageMachinelearning.**

I. INTRODUCTION

The sign Language Conversion Project is to meet thecritical requirement of providing deaf and mute people withstandardisedandaccessiblecommunication,especiallyinIndi awithitsdiverselinguisticpopulation.Signlanguageisan essential communication tool for thoughts, feelings, andexperiences. It is used in both manual and visual modalities.Though sign languages differ throughout Indian regions, thereisanurgentneedforanIndianSignLanguagestandardthatthed eafcommunityandthegeneralpubliccanbothunderstand. In the same way that most Indians find it easier tocommunicateeffectivelywhentheyspeakEnglish,the creationofanofficialIndianSignLanguagemighthelpreduce obstacles to communication and promote inclusivityacrossthe country.

TheprimaryobjectiveoftheSignLanguageconversionproject is to develop a comprehensive system that leveragescutting-edge technologies to facilitate real-time interpretationof sign language gestures. The project aims to integrate theMediapipelibraryforlandmarkdetection,enablingtheextractio nofvitalhandlandmarkinformationfromlivewebcam feeds. These landmark coordinates will serve as inputdata for a Random Forest Classifier, trained to recognize andclassify various sign language patterns with high accuracy andefficiency.Additionally,theprojectseekstoincorporatenatura llanguageprocessing(NLP)techniquestoprovideEnglish suggestions for detected signs and to translate theminto Tamil, enhancing accessibility and linguistic diversity. Bycombiningcomputervision,machinelearning,andNLPmethod ologies, the project endeavors to create a robust andversatiletoolthatempowersindividualswithhearingandspeec himpairmentstocommunicateeffectivelyandseamlessly withthebroadercommunity.

The project's ability to offer English suggestions for detectedsigns,followedbyTamilandEnglishvoicetranslations,en hancesaccessibilityandpromoteslinguisticdiversity.Beyondthes edomains,theprojecthasthepotentialforbroader societal impact, fostering greater understanding andintegrationofindividualswithhearingandspeechimpairments into mainstream social and professional spheres.As such, the scope of the project extends far beyond meretechnologicalinnovation,encompassingthepromotionofinc lusivity,accessibility,andeffectivecommunicationacross

varioussectorsofsociety.Inthispaper,Thesystemutilizesthe Mediapipe library, leveraging its landmark identificationfeatures to extract essential data from each frame of the livewebcamstream.Specifically,thecoordinatesofhandlandmark saredetectedandextracted,providingcrucialinformation about the hand gestures being performed. Theselandmark coordinates are then processed and used as input forthe subsequent steps in the translation pipeline. Once the handlandmarksaredetected,theircoordinatesarecollatedandstore d in a structured format, such as a CSV file, for furtheranalysis and training. This dataset serves as the foundation fortraining the Random Forest Classifier, a supervised learningalgorithmthatexcelsinclassificationtasks.Beforetrainin gthe model, the data may undergo preprocessing steps such asnormalizationandfeatureengineeringtoensureoptimalperform ance.

II. LITERATURESURVEY

1. GesturesinAmericanSignLanguage(ASL)arecharacterized by fast, highly articulate motion of upper body,includingarmmovementswithcomplexhandshapesandfaci al expressions. In this work, we propose a new method forword-level sign recognition from American Sign Language(ASL) using video. Our method uses both motion and handshape cues while being robust to variations of execution. Weexploit the knowledge of the body pose, estimated from anoff-the-shelf pose estimator. Using the pose as a guide, wepool spatio-temporal feature maps from different layers of a3D convolutional neural network. We train separate classifiersusing pose guided pooled features from different resolutionsandfusetheirpredictionscoresduringtesttime.Thislea dstoasignificantimprovementinperformanceontheWLASLbenc hmarkdataset[25].Theproposedapproachachieves10%, 12%, 9.5% and 6.5% performance gain on WLASL100,WLASL300, WLASL1000,WLASL2000 subsets respectively.To demonstrate the robustness of the pose guided pooling andproposed fusion mechanism, we also evaluate our method byfinetuningthemodelonanotherdataset.Thisyields10% performance improvement for the proposed method using only0.4%trainingdataduringfinetuningstage.

2. Sign language recognition (SLR) plays a crucial role inbridgingthecommunicationgapbetweenthehearingandvocally impaired community and the rest of the society. Word-level sign language recognition (WSLR) is the first importantsteptowardsunderstandingandinterpretingsignlangua ge.However, recognizing signs from videos is a challenging taskas the meaning of a word depends on a combination of subtlebodymotions,handconfigurationsandothermovements.Re cent pose-based architectures of WSLR either model boththespatialandtemporaldependenciesamongtheposesindiffer ent frames simultaneously or only model the temporalinformation without fully utilizing the spatial information. WetackletheproblemofWSLRusinganovelpose-basedapproach,whichcapturesspatialandtemporalinformationse parately and performs late fusion. Our proposed architectureexplicitly captures the spatial interactions in the video using aGraphConvolutionalNetwork(GCN).Thetemporaldependenci esbetweentheframesarecapturedusingBidirectionalEncoderRep resentationsfromTransformers(BERT). Experimental results on WLASL, a standard word-level sign language recognition dataset show that our modelsignificantlyoutperformsthestate-of-the-artonpose-basedmethodsbyachievinganimprovementinthepredictionaccur acybyupto5%.

3. Understandingcomplexhandactions,suchasassemblytasksor kitchenactivities,fromhandskeletondataisanimportant yet challenging task. In this paper, we analyze handskeleton-based complex activities by modelling dynamic handskeletons through a spatiotemporal graph convolutional neuralnetwork(ST-GCN).Thismodeljointly learnsandextractsSpatio-temporalfeaturesforactivityrecognition.Ourproposedtechnique, SymmetricSub-graphspatio-temporalgraph convolutionalneuralnetwork (S2-ST-GCN),exploitsthe symmetric nature of hand graphs to decompose them intosmallersub-graphs,whichallowustobuildaseparatetemporalmodelfortherela

tivemotionofthefingers.This

subgraphapproachcanbeimplementedefficientlybypreprocessing input data using a Haar unit based orthogonalmatrix. Then, in addition to spatial filters, separate temporalfilters can be learnedfor each sub-graph. We evaluate theperformance of the proposed method on the First-Person HandAction dataset. While the proposed method shows comparableperformance with the state of the art methods in train: test=1:1setting, it achieves this with greater stability. Furthermore, wedemonstratesignificantperformanceimprovementincomparison to state of the artmethods in thecross-personsetting, where the model did not come across a test subject'sdatawhilelearning.S2-ST-GCNalsoshowssuperiorperformance than a finger-based decomposition of the handgraph whereno preprocessingisapplied.

III.EXPERIMENTALDETAILS

AconfusionmatrixforAmericanSignLanguage(ASL)recognitionrepresentstheperformanceofaclassificationmodel by showing the predicted labels against the true labelsforeachclass(ASLlettergesture).Itprovidesdetailedinformation about the Accuracy and misclassifications of thesystem.Moreover,NaiveBayes,whichperformsbestwithindependent features, has the lowest accuracy rates. Yet, thetraitsareinterdependentinthesuggestedstrategy.

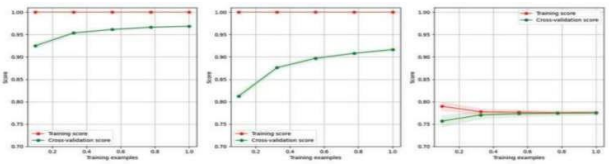
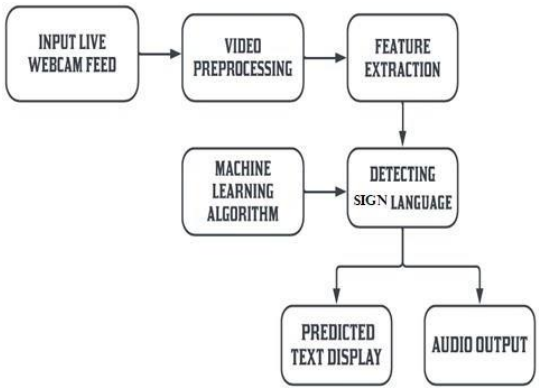


Fig- Accuracy score on the y-axis and training examples onthex-axis

The angle between landmark 12 and landmark 4, for instance,changes when the angle between landmark 0 and landmark 12(12islocatedonthetipofthemiddlefinger),whichinfluencestheothercharacteristicsaswell.Fromtheconfusionmatrix,wecancalculatevariousperformancemetrics such as accuracy, precision, recall, and F1-score. ItprovidesinsightsintowhichASLgesturesaremore challengingforthesystem torecognizeandhelpsidentifypotential areas for improvement in the recognition model ortrainingprocess.

SYSTEM ARCHITECTURE



DATAACQUISITION:

The initial or the first step of this system is vision based i.e. toacquire data at runtime via the camera. Then, these datawillbe stored in a csv file format directory, in which all the imageswill be stored and trained by the user and the saved & traineddatawillbeusedtocomparetherecentlycapturedimagetothestoreddata'sofa specificwords.

FEATUREEXTRACTION:

The palm is extracted from the data's via image segmentation.This procedure revolves around converting raw data, such asimages,intoameaningfulsetoffeaturesthatcanbeeffectivelyutilizedforanalysisandmachinelearningalgorithms. In the context of sign language recognition, theseextractedfeaturesholdvitalinformationencompassingdistinct patterns, and gestures that are indicative of variousemotional states or behaviors. To begin, the dataset undergoesessential data preprocessing steps. This involves handling anymissing data points, normalizing the data if necessary, andensuringtheoverallcleanlinessandpreparednessofthedatasetforsubsequentphases.UponloadingtheCSVfileusingrelevantprogramminglibraries,thedatarevealsitselfas

rows, each representing a sample of sign language data, whilecolumnscorrespondto specific attributes.

GESTURERECOGNITION:

Gesturerecognitionwithintherealmofsignlanguagerecognition is a critical process that involves the identificationandinterpretationofvarioushandmovementstoded ucemeaningful insights about a person's intentions, emotions, andcommunication cues. This sophisticated technology leveragesadvancementsincomputervisionandmachinelearningt otranslatephysicalgesturesintoactionableinformation.Throught heanalysisofposture,motion,andthespatialrelationships of hand sign, gesture recognition systems candiscern intricate details such as handshakes, nods, thumbs-up,andmorecomplexgestureslikepointingorevenspecificcultura lgestures.

TEXTTO SPEECH:

Once the character is successfully recognized, the resultingoutputundergoes an additional transformation from texttospeech.Thisconversionprocessisfacilitatedthroughtheutili zation of the English language process and GTTS libraryprocessing,apowerfultext-to-speechconversiontoolinPython. Unlike some other alternatives, this library operatesoffline, which ensures its compatibility and efficiency. Thisintegration enables users to observe and simultaneously hearthe translated sign language within our system, enhancing theoverallconvenienceandusabilityoftheapplication.

SUGGESTIONOFWORDS:

ThemoduledescriptionforEnglishwordsuggestionandtranslatio ntoTamilincorporatessophisticatednaturallanguageprocessing(NLP)techniquestoenhancetheaccessibilityandcomprehensionof signlanguageinterpretation. This module first utilizes NLP algorithms tosuggestEnglishwordscorrespondingtodetectedsignlanguage gestures, aiding both deaf users and non-signers inunderstandingtheintendedmessage.Leveragingadvancedlang uage models, the system generates contextually relevantword suggestions in real-time, ensuring accurate interpretationofsignlanguage cues.Additionally,themoduleintegrates

machinetranslationalgorithmstoseamlesslytranslatethesuggested English words into Tamil, catering to the linguisticdiversity of users. By providing instantaneous translations, thesystemfacilitateseffectivecommunicationandcomprehension across languagebarriers. Through thefusionof NLP technologies, this module enhances the usability andinclusivityofthesignlanguageconversionsystem,empoweringuserstoengageinseamlessandmeaningfulinteractions.

EXSITINGWORK:

BeforethedevelopmentoftheSignLanguageConversionproject,therewasnoreal-timesystemcapableofautomatically interpreting and classifying sign language cuesfrom a live webcam feed. Traditionally, understanding signlanguagerequiredhumanobservationandanalysis,which could be subjective and time-consuming. Existing computervision systems focused on basic gesture detection but lackedcomprehensive sign language interpretation. Moreover, real-time analysis of sign language using landmark detection andmachine learning was not readily available. As a result, therewas a need for an innovative system that could efficientlydetect and analyzelandmarks from live video streams, andthen classify various sign language patterns in real-time. TheSign Language Conversion project addresses these limitationsandprovidesaneffectivesolutionfornon-verbalcommunication analysis, offering significant advancements inthefieldofhuman-computerinteraction.

PROPOSEDSYSTEM:

TheproposedsystemfortheSignLanguageconversionproject integrates advanced technologies to enable real-timeinterpretationofsignlanguagegestures.UtilizingtheMediapipe library, the system detects hand landmarks fromlivewebcamfeeds,providingcrucialdataforanalysis.A

Random Forest Classifier is then employed to recognize andclassify these landmarks into various sign language patterns,ensuring accurate interpretation. Moreover, natural languageprocessingtechniquesareutilizedtogenerateEnglishsuggestions for detected signs and translate them into Tamil,enhancingaccessibilityandcomprehension.Thesystemoverlays the interpreted sign language cues, including detectedsigns and translations, onto the live video stream in real-time,facilitating immediate understanding for users. Additionally, auser-friendlyinterfaceallowsforseamlessinteractionandfeedback,ensuringongoingrefinementandimprovement.Throughrigoroustestingandvalidationprocedures,thesystem aims to deliver a reliable and effective solution forbridging communication barriers and promoting inclusivity forindividualswith hearingandspeechimpairments.

TRAINEDMODULE

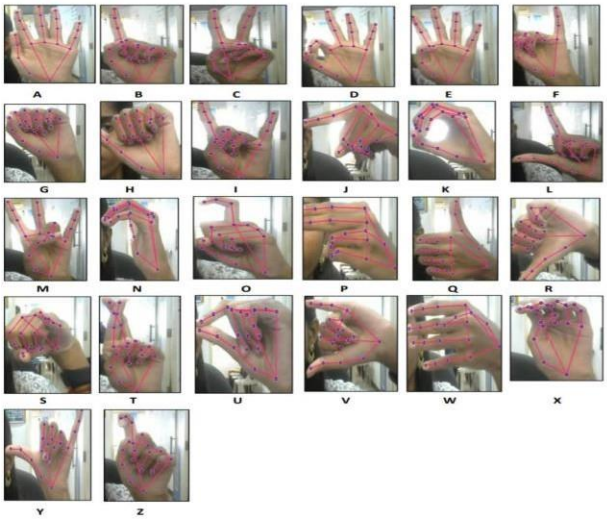


Fig-modules

III.RESULTANDDISCUSSION

Thecomprehensiveexaminationofthesimulationresultsusing the"MachineLearning"isprovidedinthispart.

additionally, employing the dataset, the suggested method'sperformanceiscontrastedwiththatofcurrentapproaches.

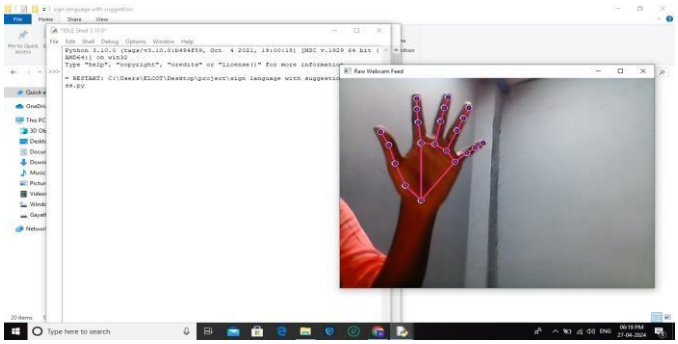


Fig.HandgestureRecognition

This method of hand identification uses the media pipe library,which is used for image processing, to first identify a handfrom an image captured by a webcam. Thus, once the hand islocated in the image, we obtain the region of interest (Roi),croptheimage,andusetheOpenCVlibrarytotransformittoa grayscale image after applying gaussianblur.TheOpenCVlibrary, commonly referred to as the Open Computer VisionLibrary, makes it simple to apply the filter. Next, we used thethresholdandadaptivethresholdmethodstoconvertthegrayscaleimagetoabinaryimage.ForthesignlettersAthroughZ,wehavegatheredpicturesofvarioussignsatvariousperspectives.

MediapipeLandmarkSystem:

Fig.LandmarkSystem

Now that we have these landmark points, we use the OpenCVlibrarytodrawthemonasimplewhitebackground.Weaddress the background and lightning situations by doing thissincethemediapipelibrary willprovideuswith landmarklocations in any backdrop and primarily under all lightningcircumstances.

Fig.LandmarkProcessingsystemusingOpenCV

V.CONCLUSION

Inconclusion,theSignLanguageConversionprojectrepresentsasignificantmilestoneinthedevelopmentofautomated, real-time systems for interpreting and classifyingsignlanguagecues.Byleveragingthepowerofcomputervision and machine learning, the project achieves remarkableaccuracyandefficiencyindetectinghandlandmarksandinterpretingnon-verbalcommunicationfromlivewebcamfeeds. The integration of a Random Forest Classifier ensuresreliableandobjectivesignlanguageclassification,contributingtothesystem'sconsistencyandeffectiveness.Moreover, the user-friendly frontend enhances the interactiveexperience, providing users with real-time analysis results andimmediatefeedback.Withdiverseapplicationsinhuman-computer interaction and user behavior analysis, the projectmarks a substantial advancement in the field of non-verbalcommunication analysis. Its success offers valuable insightsand opportunities for further research and development in thisburgeoning domain, promising to revolutionize accessibilityandinclusivityforindividualswithhearingandspeech impairments.

REFERENCES:

[1]Y. Saleh and G. F. Issa, ‘‘Arabic sign language recognitionthroughdeepneuralnetworksfine-tuning,’’Int.J.OnlineBiomed.Eng., vol.16,no. 5,pp.71–83, 2020.

[2] X.Jiang,M.Lu,andS.-H.Wang,‘‘Aneight-layerconvolutional neural network with stochastic pooling, batchnormalization and dropout for fingerspelling recognition ofChinesesignlanguage,’’MultimediaToolsAppl.,vol.79,nos.21–22,pp.15697–15715,Jun. 2020.

[3] O. Sevli and N. Kemaloğlu, ‘‘Turkish sign language digitsclassificationwith CNNusingdifferentoptimizers,’’ Int.Adv. ResearchesEng.J., vol.4, no.3, pp.200–207,Dec.2020

[3] A. Tyagi and S. Bansal, ‘‘Feature extraction technique forvision-basedIndiansignlanguagerecognitionsystem:Areview,’’ in Computational Methods and Data Engineering.Singapore:Springer, 2021, pp.39–53.

[4] M. Mukushev, ‘‘Evaluation of manual and non-manualcomponentsforsignlanguagerecognition,’’inProc.12thLang.Resour.Eval.Conf.,Eur.Lang.Resour.Assoc.(ELRA),2020, pp. 1–6.

[5] A. Tunga, S. V. Nuthalapati, and J. Wachs, ‘‘Pose-basedsign language recognition using GCN and BERT,’’ in Proc.IEEEWinterConf.Appl.Comput.Vis.Workshops(WACVW), Jan. 2021,pp.31–40.

[6] N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R.Bowden,‘‘Neuralsignlanguagetranslation,’’inProc.IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018,pp.7784–7793.

[7]S.AlyandW.Aly,‘‘DeepArSLR:Anovelsigner-independent deep learning framework for isolated Arabic signlanguagegesturesrecognition,’’IEEEAccess,vol.8,pp.83199–83212, 2020.

[8] D.Li,X.Yu,C.Xu,L.Petersson,andH.Li,‘‘Transferringcross-domainknowledgeforvideosignlanguagerecognition,’’inProc.IEEE/CVFConf.Comput.Vis.PatternRecognit.(CVPR),Jun.202

0,pp.6205–6214.

[9] R.Rastgoo,K.Kiani,andS.Escalera,‘‘Video-basedisolated hand sign language recognition using a deep cascadedmodel,’’ MultimediaTools Appl.,vol.79, nos.31–32, pp.22965–22987,Aug. 2020.