International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 24th Nov 2021. Link :http://www.ijiemr.org/downloads.php?vol=Volume-10&issue=ISSUE-11

DOI: 10.48047/IJIEMR/V10/I11/35

Title **Prototype of Educational Evaluation System Based on Speech Emotion Recognitionfor Children with SpecialEducation Needs**

Volume 10, Issue 11, Pages: 229-234

Paper Authors
Ch.HrudayaNeeharika,Y.Md.Riyazuddin

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Prototype of Educational Evaluation System Based on Speech Emotion Recognitionfor Children with SpecialEducation Needs

[1]Ch.HrudayaNeeharika,[2]Y.Md.Riyazuddin
[1]ResearchScholar,DeptofCSE,SchoolofTechnology,
GITAM(Deemed to be University),Hyderabad
[2]DeptofCSE,AssistantProfessor,SchoolofTechnology,
GITAM(DeemedtobeUniversity),Hyderabad

Abstract:

There is limited empirical research to examine whether therapeutic or educational interventions can enhance children with different developmental impairment's capacity to identify their children emotionally. Intelligent e-learning systems, speech recognition is an ever more significant field. Assisting the student's emotional side in learning activities is complex and requires a sense of the student's emotions. This paper aims to build an AI-based system to evaluate the adequate excitement level and establish a particular quantitative index for evaluation that may be utilized as a teaching assessment or teaching assistance based on pedagogical importance and detectability of emotions. This work combines the idea of local features learning blocks (LFLBs) for extracting the features with a parallel block of CNNs with a range of filter longitudes for collecting multi-temporal data. The proposed affective arousal teaching system may simultaneously do process assessment in class. Results indicated that the emotion identification training provided in an intervention program based on conduct could significantly increase children's emotional recognition at a wide range of abilities. The results suggest that the proposed architecture may deliver similar outcomes at the advanced level despite data increases and advanced pre-processing.

Keywords: Feature Learning Speech Emotion Recognition, children with special education needs, Affective arousal

Introduction

The future of schooling is integrally related to new technological advancements and new intelligent machine computercapabilities. In this sector, progress in artificial intelligence is open to new educational and higher education challengesandopportunities,withpotentialforsignificantchangesingovernanceand highereducationinstitutions'internalarchitecture. Childrenmustbe ready forfuture economies'productive contributionand future societies to becomeresponsibleandengagedcitizens[1,2].Furthermore,artificialintelligence(AI)improvesinstruments andtoolsuseddailyin citiesandcampusesworldwide.Websearch engines,smartphonesandapplications,publictransitandhomeappliances. For example, Siri is a classic example of artificialintelligence,asophisticatedcollectionofcontributionstotheprojectandsoftwaresolutionsthathavebeenincludedindailylife[3,4].

Disabled people are also known as special needs people [5]. The term special needs have been widely used in the last several years as a synonymfor disability. Standard techniques for can successfully enhance the emotional understanding of learners [6], intellectual handicaps metadata extraction rely on the video's visualinformation. However, the content delivered consists not just of visual information but also auditory information helpfultodeterminecontext,emotions,andothercontentmetadata.Consequently,itisrelevanttorecognizeemotion fromspeechwhile producing metadata and using all available content data. The primary aim is to give an advantage to students andteachers compared with techniques that do not use technology. It might be challenging to incorporate instructionaltechnology into an educational environment. The integration process should take into account problems that must be addressed in a particular students' class.

Technology can help manage unique educational challenges or infrastructure for non-technological activities that have not been implemented. While studies such as these show that emotional comprehension and the recognition of emotions, in general, are essential developmental factors, there is less evidence on the efficacy of efforts to modifyemotional awareness in persons with impairments. Adult research has usuallyshown that treatments

[7], and functions with high autism [8], or brain injuries [9]. However, child-centered research was

less consistent. For instance, [10] found no increase in deaf children'semotional detection skills in an eleven-lesson psych educational program.

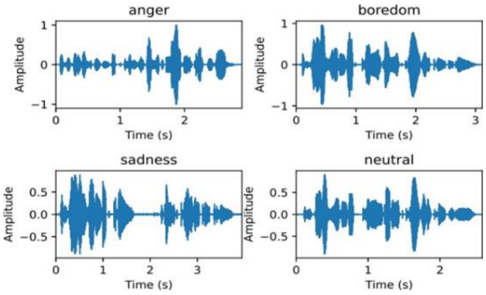


Figure 1: Raw waveform plots using the same sentence and speaker, different portrayed emotions

Most of the studies in emotional recognition of speech included the technique of collecting vocal emotions acoustic characteristics. They based this onthenotionthatvariouswaveformcharacteristicsmay assesschangesinspeechproducedby different arousal or valence conditions in the speaker, as illustrated in Figure 1. They provide a proposal for a specific technique, depending on the emotion recognizedandthelong-termattitudeofthechild.Whiletheyhaveconducteddetailedstudies on facial expression identification, few empirical studies have carried out facial expression in e-learning systems.Thepresenceintheeducationofemotions remains unadjusted.Inrecentyears,DeepLearning(DL),which hasoutperformed classic approaches using neural network topologies like CNN, and different recurrent neural network operations, has emerged (RNNs). DL with CNNshasalsoallowednetworkstoimmediatelylearnandextractfeaturesfromthe raw audio input, eliminating the need for complex feature engineering manually. This study will examine the use ofraw audio waveforms to combine parallel CNNs to extract features and long-term memory networks (LSTMs) to classifyspeechemotionaldetectiontasks(SER).

This study examined if emotion recognition is an emotional attribute that may be controlled through a behaviorallybased evaluation and procedure. We assumed young children subjected to direct education in emotional awareness withdevelopmental delays and disabilities would show considerable progress in their capacity to understand both basic andadvancedemotions.

Related work:

Using self-intelligence models and speech recognition are essential elements of their critical study in creating helpapplications appropriate for children with cognitive impairments. But they have realized the highest progress of mobiletechnologyinrecentyears.Theyrepresentsignificanttechnologicaldevelopmentsandarefrequentlythemoststraightforwardcomputertechnologyintheworld.

SpecAugment, a simple approach for increasing speech recognition, was presented in [11] by the authors. The feature inputs on the neural network(e.g.,bankcoefficientsoffilter)areimmediatelyappliedforSpecAugment.Thepolicyofincreaseistodistortion features,maskfrequencychannelblocksor masktimeblocks.For end-to-endvoicerecognitiontasks, we use Spec Augment on Listen, Attend, and Spell networks. The 300h hands of the LibriSpeech 960 Switchboard achieved state-of-the-artperformance,overcomingallprevioustasks.OnLibriSpeech,6.8%WERwithoutlanguagemodelin a test-other, and 5.8% WER with an acceptable language model in a test-other. They compared this with the current7.5% WERhybridsystem.ForSwitchboard,theHub5'00Testsareachievedat7.2%/14.6%ontheHub5'00 Switchboard/CallHome partwithoutusingalanguagemodelandat6.8%/14.1% onlowfusion,comparedtothepriorstate-of-the-arthybridsystemat8.3%/17.3%WER.

Using a generic model is recognized as the standard approach for speech emotion recognition emotions based upon the voices of different persons. These approaches cannot consider the specific type of customized communication. The recognized outcomes, therefore, range significantly from each individual. Authors in [12] suggested an adaptive emotion recognition framework using user instant feedback data that would create a personal adaptive recognitionmodel by prompting labeling approach, which could be applied to each user in a mobile device setting. They may recognize emotions through the construction of a customized model—the suggested framework. The frameworksuggested was assessed in three comparison experiments to be better thanstandard research approaches. The paradigm suggested can be used in healthcare, emotion surveillance, and individual services.Regrettably, the present speech enhancement modulation approaches produce limited performance in detecting stressful human emotions when noise is unavoidable and changes every vehicle position. In this respect, they suggest front-end processingframes in various non-stationary noisy settings, particularly for stress emotion detection instances. Thisstudy [13] covers three interrelated issues: the assessment, modification, and synthesis of noisy speech in real-time background noises, extraction from the noisy voice stimuli tospeechemotions,andsystemperformanceevalu ationthroughobjectiveparametersandconfusion matrix.

Theauthorssuggestedanactivegrouplearningfunctionalmethodin[14]thatreducesthemis-conformitiesbetweentraincircumstancesandtestconditionsandprovidesdifferentclassificationswithintheensemble.Theresultsshowed that selecting features in a small group from the target domain can yield significant improvements. The technique suggested also showed the significance of choosing samples for

annotationusingthepropercriterion,wherevotingentropyispreferrediftheselectedsamplesizeissmall.Randomsamplingistheidealapproachwheneverthesamplesizerisesbecause the distribution of the target domain is better represented. They implemented the system with a set of SVMs. Theadvantagesoftheexperimentalevaluationforotherclassifierslikerandomforestareinterestedinexploring.Theyalso intended to assess other AIcriteriaforfunctionselection,whichtakethedatadistributionandtheuncertaintyintoaccount.

The authorsinvestigatedthenetworksofscientificcooperationbetweenspecializededucationandspeechtherapy in [15]. The corpus ofthisstudycomprises267paperspublishedby44scholarswhosedissertationsandthesescharacterize the intersection between these fieldsofknowledge, whichcompletedpostgraduatestudiesattheFederalUniversityofSãoCarlos between 1981 and 2010. Lattes' curriculum was the source of the data. The approach used was a Social Network Analysis (SNA) designed to develop scientific working connectionsamongstplayersengagedinvariousknowledgesectorsthroughthecreationandco-authoringofthenetworks.UcinetandNetdrawtoolshavebeenusedtomapandcreategraphsfor actor cooperation. Results revealed smaller clusters with few participants in the publication field; the creation inpartnership with scholars in the nation and abroad of collaborative networks between advisors and student publications.The study also showed that examining the Special Education and Speech Therapy scientific collapsing networks helpsbuildfutureresearchonthisinterface.

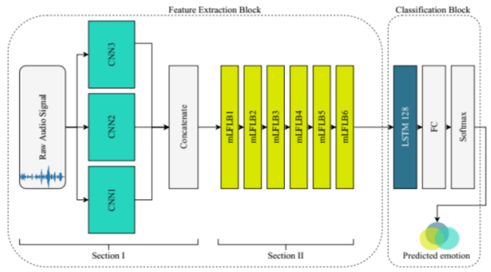
Material and Methods

This sectionoutlines what the recommended network architecture looks like one feature extractionblock and oneclassification block. Figure 2 then shows the resources and data sets utilized to train the network. The extraction featuresblock comprises parallel convolutional layers which extract three different temporal resolutions from the speech, then arecombinedandtransformedintoaseriesofLFLBs,which extractstheessentialfeaturesanddecreasestheresolution oftheclassification block representation. The classification block comprises an LSTM layer, a fully connected layer (FC), and a layer of Softmax, which generates a final classification viewofthenetwork'soutputs.Inbothblocks,learningis optimized.

Figure 2:Proposed architecture with parallel multi-temporal convolutionallayersandaseriesofmodifiedLFLBs

Feature Extraction Block:

The extraction block of features is vital for the learning of raw signal features. Features that give predictive value to themodelcontributetotheaccurateclassificationofunseendata.At16kHz,araw128000-



bitvectorisusedtoshowtherawinput audio signal. This audio vector has to be reduced to dimensionality for the classification block and LSTM to learneffectively.Thestepsorpoolingcanlowerasignal'sdimensionalityaspartofthefeatureextraction.

Classification Block:

The classification block is relatively straightforward,comprisinganLSTMlayer,acompletelylinkedSoftmaxlayer.Based on much prior research, wehaveestablishedauni-directionalLSTMunitthatwillcontributelittletonetworkperformanceforthefuture.Wemaychange andchecktheaccuracyofthecells intheLSTM, testing64,128, and256.

Dataset:

This study's dataset is a language database [16] that comprises two-child speech recordings of various speaking activities. The first (healthy) subgroup contains recordings ofchildrenwithoutspeechproblemsandthesecond(patients)SLI-relatedchildren. The severity of these children is variable (1–mild, 2–moderate, and 3–severe). They recorded the corpus in aschoolroom and a clinic in the natural setting. 44 Native Czech members (15 boys and 29 girls) aged between 4 and 12 years of age were registered in this subgroupover2003-2005(inFrench).Aprivatespeechtherapistpracticehas registereda database of children with specific language impairment (SLI). There are two components in the database. The first partis the database recording. In the background's presence of

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

noise, someone typically established these databases in aschoolroomorthe consultation roomof aspeechand language therapist. This setting replicates children's naturalsurroundings and is essential for recording children's usual conduct. Additional recordings of specific children are part ofthesecondcomponent.

Pre-Processing:
We aim to reduce the preprocessing section to identify to what extent extraction features we may leave to the model. Togenerate the model training data, the 16 kHz sampling rate of the Nyquist-Shannon theorem enables us to evaluateinformation withoutobjectsatfrequenciesofupto8kHz,themaximumfrequencyofordinaryhumanlanguage.Wehavea one-dimensional floating-point vector after sampling the audio stream. There might be different volumes for
each audiofile.Thus,regardingtheroot-mean-square(RMS),westandardizethesignalvalues(amplitudes).Wehaveappliednodataaugmentation to any dataset. Data growth adds complexity, and this study aims to carry out minimum pre-processingmanually;hence,dataincreases werechosennottobeincludedbythisstudy.
each audiofile.Thus,regardingtheroot-mean-square(RMS),westandardizethesignalvalues(amplitudes).Wehaveappliednodataaugmentation to any dataset. Data growth adds complexity, and this study aims to carry out minimum pre-processingmanually;hence,dataincreases werechosennottobeincludedbythisstudy.

Results and Discussion

Semantic featuresandauditoryfeaturesareincluded.Extractingacousticfeaturesthatarebasicandadequatetoaccomplishthe classification effect is used to implement speech recognition emotion throughout the teaching process. We convertedvideo data into emotional multiple time series using the procedure mentioned above. The output layer for the system is intheassessmentindexdesignmodule.WealsoincludedtheMFCCclassification andtheoptimizedMFCCclassificationinTable1-2.

Methods	Anger	fear	happy	neutral	sad	surprise
Traditional method	69.21%	78.94%	64.42%	89.21%	91.20%	92.31%
Proposed method	74.52%	81.28%	69.85%	91.20%	93.21%	94.52%

Table 1: Result of MFCC classification

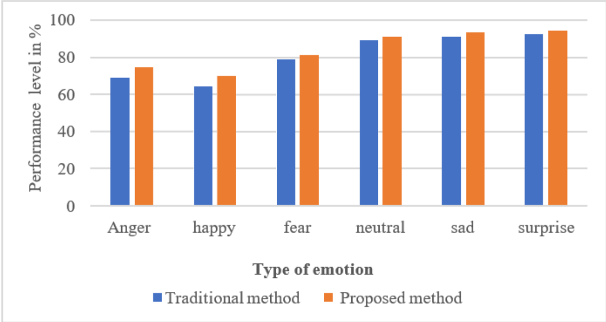


Figure3:ComparisonresultoftwomethodsMFCCclassification

Figure 3 shows the average classification accuracy of 6 emotions common MFCC features for the proposed method is84.10%, outperformsasrelatedtothetraditionalmethodis80.80 %.

Methods	Anger	fear	happy	neutral	sad	surprise
Traditional method	71.21%	79.24%	81.20%	90.24%	92.31%	93.45%
Proposed method	75.26%	84.52%	86.23%	92.51%	94.58%	95.84%

Table2:ResultofoptimizedMFCCclassification

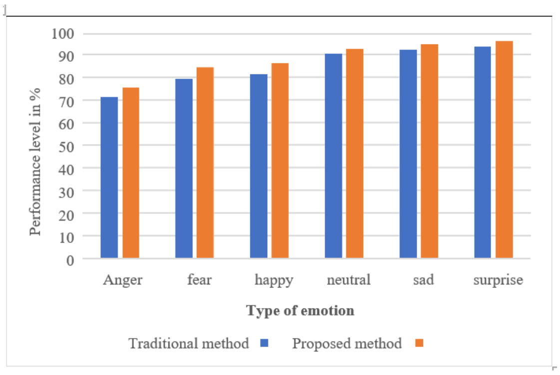


Figure4:ComparisonresultoftwomethodsoptimizedMFCC classification

Figure 4 shows the average classification accuracy of 6 emotions common MFCC features for the proposed method is89.16%, outperformsasrelatedtothetraditional methodis84.65%.

With thetrainingandevaluation ofthedataset,our LSTMarchitectureproducedthefollowingdata.Theimproveddesignwasimplementedstraightwithoutfurthertweakingusingdata setvalidationdata,enablingthedata setvalidationfindingstobepuretestresults.Table3showsthemaximumprecisiononeachfoldfor supportsizefor each emotionalclassineachfold.Table3showsthethreefold cross-validation.

Details of folds	Traditional method Accuracy in (%)	Proposed method Accuracy in (%)
First fold	85.32	91.24
Second fold	86.41	92.31
Third fold	88.92	93.74

Table3:Speaker-dependentaccuracyforeachofthethreefolds,validating proposed modelonadataset

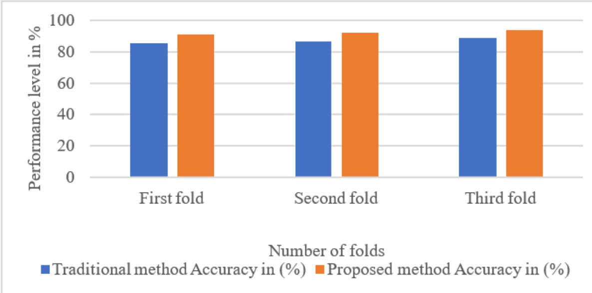


Figure5:Comparisonresultoftwomethodsforperformancel evelversus thenumberoffolds

Figure 5 shows the average classification accuracy for the proposed method is 93.74 % for three folds outperforms asrelatedtothetraditionalmethodis88.92%.

Conclusions

The study examined how AI Technologies affect every child's life and help children with special needs to live moreefficiently. AI software will substitute many activities at the heart of higher education instruction based on complicated algorithms developed by programmers who may communicate their priorities or agendas on operating systems. And we experimented with our suggested system with pupils aged 8 to 12 years. The findings reveal that emotions have beenidentified,andthesystemhasbeenup-to-date.Thisstudyalsooffersanovelprocess-basedautomaticassessmentmethod,entirelydifferentfromtheoldtechnique, byresearchingthesubjectiveimpressionsofstudentsorthetestresultsaspartoftheassessmentofteachingquality.Basedon severaltestinganderror techniques,tweaking,etc.,themodelwasacomplexeffort. We highly trained the model for differentiating between the voices of the boy and the girl and determines 98% accuracy. The model detected emotions with much over 92%accuracy.Moreaudiofilesfortrainingcanenhance accuracy.

References

[1] Karadimou, Maria & Tsioumis, Konstantinos. (2021). Preparing kindergarten students for future active citizens. Journal of Studies in Education. 11. 23. https://doi.org/10.5296/jse.v11i3.18693.

[2] Богославець, Л & Кікена, Г. (2018). FORMATION OF LINGUISTIC COMPETENCE OF FUTURE ECONOMIES IN PROFESSIONAL PREPARATION. Proceedings of the National Aviation University. Series: Pedagogy, Psychology. https://doi.org/10.18372/2411-264X.12.12919.

[3] Dumitriu, Dan & Popescu, Mirona. (2020). Artificial Intelligence Solutions for Digital Marketing. Procedia Manufacturing. 46. 630-636. https://doi.org/10.1016/j.promfg.2020.03.090.

[4] Meriyem, Chergui & Aziza, Chakir. (2020). IT Governance Knowledge: From Repositories to Artificial Intelligence Solutions. Journal of Engineering Science and Technology Review. 13. https://doi.org/10.25103/jestr.135.09.

[5] Mickahail, Bethany & Andrews, Kate. (2018). Embracing People with Special Needs and Disabilities. https://doi.org/10.1007/978-3-319-54993-4_8.

[6] Seo, W.S. & Jeong, Y.C. & Sea, H. (2012). Deficits of emotional recognition ability in ADHD children. Neuropsychiatrie de

[7] l'Enfance et de l'Adolescence. 60. S264-S265. https://doi.org/10.1016/j.neurenf.2012.04.693.

[8] Zaja, Rebecca & Rojahn, Johannes. (2008). Facial emotion recognition in intellectual disabilities. Current opinion in psychiatry. 21. 441-4.https://doi.org/10.1097/YCO.0b013e328305e5fd.Yeun

[9] g, Michael & Chan, Agnes. (2020). Executive function, motivation, and emotion recognition in high-functioning autism spectrum disorder. Research in Developmental Disabilities. 105. https://doi.org/10.1016/j.ridd.2020.103730.

[10] Jorna, Lieke & Westerhof-Evers, Herma & Khosdelazad, Sara & Rakers, S.E. & Naalt, Joukje & Groen, Rob & Buunk, Anne & Spikman, Jacoba. (2021). Behaviors of Concern after Acquired Brain Injury: The Role of Negative Emotion Recognition and Anger Misattribution. Journal of the International Neuropsychological Society. 1-9. https://doi.org/10.1017/S135561772000140X.

[11] Dyck, M. J., & Denver, E. (2003). Can the emotion recognition ability of deaf children be enhanced? A pilot study. Journal of Deaf Studies and Deaf Education, 8, 348-356.

[12] Park, Daniel & Chan, William & Zhang, Yu & Chiu, Chung-Cheng & Zoph, Barret & Cubuk, Ekin & Le, Quoc. (2019). SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition. 2613-2617. https://doi.org/10.21437/Interspeech.2019-2680.

[13] Bang, Jae & Lee, Sungyoung. (2015). Adaptive Speech Emotion Recognition Framework Using Prompted Labeling Technique. KIISE Transactions on Computing Practices. 21. 160-165. https://doi.org/10.5626/KTCP.2015.21.2.160.

[14] Paikrao, Pavan & Mukherjee, Amrit & Jain, Deepak & Chatterjee, Pushpita & Alnumay, Waleed. (2021). Smart emotion recognition framework: A secured IOVT

[15] perspective. IEEE Consumer Electronics Magazine. PP. 1-1. https://doi.org/10.1109/MCE.2021.3062802.

[16] Akhiat, Yassine & Chahhou, Mohamed&

Zinedine, Ahmed. (2019). Ensemble Feature Selection Algorithm. International Journal of Intelligent Systems and Applications. 11. 24-31. https://doi.org/10.5815/ijisa.2019.01.03.

[17] Hayashi, Carlos & Hayashi, Maria Cristina. (2012). Analysis of scientific colaboration networks between special education and speech therapy. Revista Interamericana de Bibliotecología. 35. 285-397.

[18] Grill P, Tučková J (2016) Speech Databases of Typical Children and Children with SLI. PLoS ONE 11(3): e0150365. https://doi.org/10.1371/journal.pone.0150365.