iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: X Month of publication: October 2025

DOI: https://doi.org/10.22214/ijraset.2025.74862

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

An Investigation of PLP Variants for Acoustic Scene Classification

Chandrasekhar Paseddula
Electronics and Comunication Engineering G.Narayanamma Institute of Technology and Science, Hyderabad, Telangana, India

Abstract:Inthispaper,PLPcoefficientsandPLPCCfeaturesareinvestigatedasarepresentationofanacousticsceneusingDNN.Wehaveexper- imentedonDCASE2018Task1datasetandDCASE2017dataset. Experiments are carried out for subtasks A and B. We have exper- imented with individual feature sets as well as decision level DNN scorefusionsofdifferentcombinationsoffeaturesets. Fromtheex- periments, it was observed that the proposed PLP and PLPCC give betterperformanceforsubtasksAandB. ForsubtasksAandB,in- dividualPLPyieldanimprovementof8.9%and13.6%respectively. FurtherPLPCCresultedinanimprovement of8.6% and12.5%. We haveachievedsignificantimprovementsinaccuracyforsubtasksA (11.4%)andB(14.4%)afterfusionofDNNdecisionlevelscoresob- tainedfromPLP,PLPCCandlogmel-bandenergiescomparedtothe 2018 baseline system.We have also experimented on 2017 dataset on4foldcross-validation,withindividualPLPyieldinganimprove- ment of 5.8% and PLPCC achieving an improvement of 4.7%.The fusionofDNNdecisionlevelscoresobtainedfromPLP,PLPCCand log mel-band energies gave an improvement of 6.0% compared to the 2017 baseline system.

Index Terms: Log-Melbandenergies,PerceptualLinearPrediction (PLP),AcousticSceneClassification(ASC),DeepNeuralNetwork (DNN).

I. INTRODUCTION

The Acoustic Scene Classification (ASC) research in signal pro- cessing, machine learning and interdisciplinary fields has become more popular due to significance of information gathered from en- vironmentalsoundsinvariousapplicationslikesurveillance,smart- phones, robotics, data archving, audio hearing aids, etc [2,3].In- tially, a large number of spectral, cepstral, energy and voicing- relatedaudiofeaturesandSVMareusedtoclassifytheseshortseg- ments and a majority voting scheme is employed in [4].Spectral, temporal and spatial features with SVM classifier is used for ASC in [5].Histogram of gradients of time-frequency representationsfor audio scene detection is investigated in [6].A Bag-of-Features approach is used for acoustic event detection in [7].Spectrogram pattern matching based environmental sound classification is used in [8].Deep convolutional neural networks and data augmentation for environmental sound classification is carried out in [9].Sound scene identification based on MFCC, binaural features and a sup- portvectormachineclassifierisusedforASCin[10]. Framebased classificationusingHiddenMarkovmodelwasproposedin[11].A hybridaproachusingbinauralI-vectorsanddeepconvolutionalneu- ralnetworksisusedforASCin[12].ASCusingaCNN-superVector systemtrainedwithauditoryandspectrogramimagefeaturesisused in [13].Generative adversarial network based acoustic scene train- ing set augmentation and selection using SVM hyper-plane is pro- posedin[14].Anaudiotrackrepresentedbylongtermstatistical distributionofsomeshorttermspectralfeatures(melfrequencycep- stral coefficients (MFCC)) is used for ASC in [15].Acoustic scene classification using matrix factorization with unsupervised feature learningiscarriedoutin[16]. Acompactanddiscriminativefeature based on auditory summary statistics for ASC is proposed in [17]. Bag of Visual Words (BoVW) representations are used for ASC in [18].Waveletsarerevisitedfortheclassificationofacousticscenes in [19,20].Wavelet transform based mel-scaled features are used forASCin[21].Ensembleofspectrogramsbasedonadaptivetem- poral divisions for ASC is used in [22].Fully CNNs and I-Vectors were proposed in [23].X-Vector embedding and CNN model for ASC were proposed in [24].From the current research in the field of deep learing for all machine learing applications, and its succees in DCASE 2016, DCASE 2017 [25] and DCASE 2018 [1].

Acoustic scenes datasets captured from the sorrounding environ- ments cover the audio frequency range from 20Hz to 20kHz.Fea- tures that can capture local information in both time and frequency domain would provide good representation of acoustic scenes.The results from previous research on DCASE 2013, 2016, 2017, and 2018 and analyses show the need for a suitable feature-classifier combination. Also,theconsequentacousticscenesconstituteavari- etyofenvironmentalsoundsthatcanformcomplexsounds. There- fore, only one particular type of feature may not be sufficient to effectively and discriminatively represent them.In this paper, we propose the use of PLP and its variants for ASC using DNN as a classifier and also DNN score level fusion for decision.

Thepaperisorganizedasfollows: Section2describesdetailsof featureextraction,Section3describestheproposedmethod.Section4 discussestheresultsandanalysisoftheproposedsystem.Section5 providesthesummaryandconclusions.

II. FEATURE EXTRACTION

In the present work, we are investigating PLP and PLPCC features for ASC. The description of these features are given bellow.

A. Perceptual Linear Prediction

The perceptual linear prediction model was developed by Herman- sky[26]. PLPwasproposedtomatchcharacteristicsofhumanaudi- torysystem. Fig. 1. showstheessientialcomponentsofPLPfeature extraction.PLP has three important perceptual aspects: first- the critical-band resolution analysis, second- the equal-loudness analy- sis, and third- intensity-loudness conversion (cubic-root).Intially, the windowed signal power spectrum is calculated as,

$$P(\omega)=Re[S(\omega)]^2+Im[S(\omega)]^2 \qquad (1)$$

The first component is a conversion from audio frequency to Bark scalefrequency,Barkfrequencyagoodrepresentationofthehuman hearingresolutioninfrequency.Thebarkfrequencyequivalenttoan

Table 1. Acoustic scene classification results for DCASE 2018 task1 subtask A (P1: PLP with DNN, P2: PLPCC with DNN, P3: Log-Mel withDNN,P4: DNNscorelevelfusionofPLPandPLPCC,P5: DNNscorelevelfusionofPLPCCandLog-Mel,P6: DNNscorelevelfusion of PLP and Log-Mel, P7: DNN score level fusion of PLP, PLPCC and Log-Mel).

AcousticScene(%)	Baseline-2018[1]	P1	P2	P3	P4	P5	P6	P7
Airport	72.9	71.3	66.0	55.5	69.1	67.5	70.6	69.8
Bus	62.9	72.7	71.9	72.3	73.6	72.7	74.8	74.4
Metro	51.2	70.1	69.0	62.8	71.6	73.6	70.5	72.4
Metrostation	55.4	65.6	65.3	63.3	66.8	65.3	65.3	64.5
Park	79.1	84.7	83.5	81.0	84.3	83.9	84.7	84.7
Publicsquare	40.4	54.6	50.9	51.9	55.1	47.7	52.8	50.5
Shoppingmall	49.6	49.1	59.9	83.5	52.3	78.9	74.9	77.8
Streetpedistrain	50.0	57.9	59.5	52.6	58.3	56.3	55.5	55.9
Streettraffic	80.5	89.0	89.0	89.4	89.0	89.0	88.6	88.6
Tram	55.0	71.3	68.2	70.5	69.3	72.0	72.0	72.0
Average	59.7(±0.7)	68.6	68.3	68.3	68.9	70.7	71.0	**71.1**

Table2.Acousticsceneclassificationresultsfortask1subtaskBfortheproposedsystem(P7).

AcousticScene(%)	Baseline-2018[1]				P7			
	A	B	C	Average(B,C)	A	B	C	Average(B,C)
Airport	73.4	68.9	76.1	72.5	61.9	27.8	44.4	36.1
Bus	56.7	70.6	86.1	78.3	71.1	83.3	83.3	83.3
Metro	46.6	23.9	17.2	20.6	66.3	38.9	72.2	55.5
MetroStation	52.9	33.9	31.7	32.8	61.4	38.9	38.9	38.9
Park	80.8	67.2	51.1	59.1	84.7	94.4	100.0	97.2
Publicsquare	37.9	22.8	26.7	24.7	48.6	38.9	50.0	44.4
Shoppingmall	46.4	58.3	63.9	61.1	74.6	88.9	77.8	83.3
Streetpedestrien	55.5	16.7	25.0	20.8	62.3	33.3	44.4	38.8
Streettraffic	82.5	69.4	63.3	66.3	86.2	77.8	83.3	80.5
Tram	56.5	18.9	20.6	19.7	74.7	38.9	44.4	41.6
Average	58.9(±0.8)	45.1(±3.6)	46.2(±4.2)	45.6(±3.6)	69.2	56.1	63.9	**60.0**

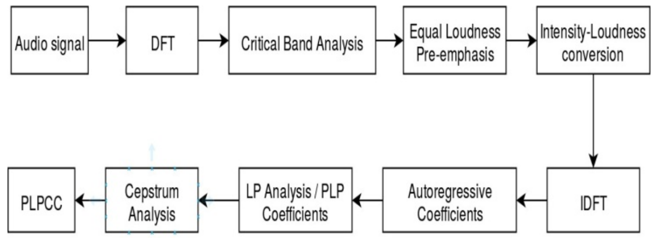


Fig. 1.Block diagram of the PLP cepstral coefficients feature ex- traction.

audiofrequencyis,

$$\Omega(\omega)=6 ln\omega/1200\pi+ [(\omega/1200\pi)^2+ 1]^{0.5}\}. \qquad (2)$$

Secondcomponent,theauditorypowerspectrumisconvolutedwith the power spectrum of the critical-band masking curve to simu-latethehumanhearingcritical-bandintegration.Finally, smoothed powerspectrumisdownsampledatintervalsof≈1-Barkintervals. Thethreecomponentsfrequencywarping,smoothingandsampling are integrated into a one filter-bank called Bark filter bank.An equal-loudness pre-emphasis weighs filter-bank results to simulate thehearingsensitivity. Theequalizedvaluesarechangedaccording tothepowerlawofStevensbyraisingeachtothepowerof0.33.

The yielding auditory warped power spectrum is then processed by linear prediction (LP). Applying LP to the auditory warped power spectrum means, compute the predictor coefficients of a signal that has warped spectrum as a power spectrum and the predictor coeffi- cients are termed as PLP coefficients.Finally, cepstral coefficients are achieved from the predictor coefficients by a recursion that is equivalent to the logarithm of the model spectrum followed by an inverse Discrete Fourier transform (IDFT) . These coefficients are termed as PLP cepstral coefficients (PLPCC) [27].

Thefeaturesetisextractedfor40filterbankswithaframesize of 40 ms with 20 ms overlap.Even for stacked context of features, static PLP (40 dimensions), Δ PLP (40 dimensions) and ΔΔ PLP (40dimensions)featurevectorsare extractedmakingit120dimen-sionsperframe. ThesameapproachisfollowedforPLPCCfeature extraction.

III. PROPOSED SYSTEM

TheproposedASCsystemincorporatesthegeneralASCframework asshowninFig. 2. Fromdevelopementdata, allaudiosignalspass through pre-processing and feature extraction processes.Since the DCASE data is in binaural stereo format, the first pre-processing stepistoconvertthedatasamplestomonophonicaudiobyaverag- ing the two channels.Pre-emphasis factor of 0.97 was used to em-phasizethehigh-frequencycontent. Thefeatures(PLP,PLPCCand Log-Mel band energies) are extracted from the preprocessed data. Modelswerebuiltfromfeaturesoftrainingdataandthenemployed forclassificationofthetestsamplesusingDNNmodel.

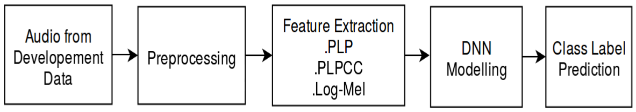


Fig.2.Blockdiagramofproposedsystemforindividulfeatures.

A. DescriptionoftheProposedSystem

Intheproposedapproach,wehaveconsideredDNNmodelasitwill have fewer computations and also less training parameters than the baselinemodelCNN[1].FromthissectiononwardsLog-Melband energiesisnotedasLog-Mel.Wehaveexperimentedwithindividual feature sets and combinations of DNN scores have been obtained with each feature set.The following are the seven set of proposed exeperiments:

P1: PLP with DNN, P2: PLPCC with DNN, P3: Log-Mel with DNN, P4:DNN score level fusion of PLP and PLPCC, P5:DNN score level fusion of PLPCC and Log-Mel, P6:DNN score level fusion of PLP and Log-Mel, P7:DNN score level fusion of PLP, PLPCC and Log-Mel

B. DetailsofDNNClassifier

TheDNNusedinthisstudyisafullyconnectedfeedforwardneural network.The network has an input layer, three hidden layers, and anoutputlayer.

Theinputlayerhas120(dimensionoffeaturevec- tor)neuronswithlinearactivationfunctionwhileeachhiddenlayer has 200 neurons with rectified linear unit (ReLU) activation func- tion. Theoutputlayerhasneurons(numberofclasses)withsoftmax activationandcategoricalcross-entropylossfunction.ADAM[28] optimizerisusedhereforbetterweightoptimizationanditslearning rateissetto0.005. L2regularizerisusedtoavoidoverfittingwitha value of 0.000099 and DNN trained with 30 epochs.

C. Decision Strategy

For individual feature sets, we used Max rule for classification. Computation of DNN score fusion of any two different features is doneasfollows:LetusconsiderthefusionofPLPandLog-Mel bandenergies.IfX^{i}_{PLP} and$X^{iLog-Melbandenergies}$aretheDNN scores generated by two models for the ithacoustic scene, then a combined score is given by

$$X_d^{icombined} = \alpha X_P^{iPL} + (1-\alpha) X^{iLog-Melbandenergies} \quad . \tag{3}$$

Similar procedure is carried out to compute DNN score fusion for different combinations of two feature sets.Computation of DNN score fusion of three different features is given by

$$X_d^{icombine} = \alpha X_P^{iPL} + \beta X_C^{iPLPC} + \gamma X^{iLog-Melbandenergies} \tag{4}$$

where,summationof α,βandγisone.

In theseproposedsystems,forP1 toP6,α=0.5andfor P7, α= 0.5,β=0.2 andγ= 0.3 isfixedtoobtainsignificantim- provements.

where,summationof α,βandγisone.
In theseproposedsystems,forP1 toP6,α=0.5andfor P7,
α= 0.5,β=0.2 andγ= 0.3 isfixedtoobtainsignificantim- provements.

IV. RESULTS AND DISCUSSIONS

This section gives the datasets used, baseline methods, results and discussionsonASCtasksforDCASE2018Task1subtaskA(basic ASC)andsubtaskB(ASCwithmismatchedrecordingdevices)and DCASE 2017 (ASC). The results of DCASE 2018 task1 subtask A are presented in Table 1 and results of subtask B are presented in Tables2and3respectively.TheresultsofDCASE2017ASCTask are presented in Tables 4 and 5.

A. Datasets and Baseline Methods

We have used the development dataset of TUT Acoustic Scenes 2018 [1] and TUT Acoustic Scenes 2017 [25].According to the DCASE2017challenges'ASCtasksetup,developmentdataispar- titioned into k folds, where k=4 for both the datasets.Fold-wise meanclassificationaccuracyisusedastheperformancemetricdur- ing development.For performance comparison, we have used the baselinesystemsoftheDCASEchallengesof2018and2017,which are Log-Mel band energies with CNN [1] and Log-Mel band ener- gies with MLP [25].

B. AnalysisofResultsonDCASE2018SubtaskA

The results for the subtask A are given in Table 1 for the DCASE 2018 baseline system and proposed systems (P1-P7).This subtask is concerned with the basic problem of ASC, in which all available dataisrecordedwiththesamedevice,whichinthiscaseisdeviceA (Zoom F8 audio recorder).

From the table, it can be observed that individual PLP features performbetterthanPLPCCandLog-Melfeatures. Fortwofeatures combination proposed systems (P4-P6), performance has increased consistently, out of which P6 has given better performance than P4 and P5 with a relative performance of 11.3% as compared to base- line system.

Further, it is observed that the DNN score fusion of PLP, PLPCC and Log-Mel band energies features (P7) has given significant improvement in accuracy and indicates complementary informationbasedon11.4%ofrelativeimprovement. Usingprosed features (P7) the Shopping mall, Park and Street traffic classes are wellclassifiedwhencomparedtootherclassesforsubtaskA.From thetable,itcanalsobeobservedthattheproposedsystemhasgiven animprovementintheaverageaccuracy.Therelativeimprovements of8.9%,8.6%,8.6%,9.2%,11.0%,11.3%,and11.4%areobtained for P1-P7 respectively as compared to the DCASE 2018 baseline system.

C. AnalysisofResultsonDCASE2018SubtaskB

TheresultsforthesubtaskBarepresentedinTables2and3.This subtask is concerned with the situation in which an application will betestedwith afew differenttypes ofdevices (deviceA, deviceB- Samsung Galaxy S7 and device C-IPhone SE), preferably not the samedeviceastheonesusedtorecordthedevelopmentdata.Theresults of the Subtask B are given in Table 2 for the DCASE 2018 baselineandtheproposedsystemwhichusestheDNNscorefusion of PLP, PLPCC and Log-Mel band energies features (P7). Fromthistable,itcanbenotedthatindividualPLPfeaturesper- formbetterthanPLPCCandLog-Melfeatures.Fortwofeatures combination proposed systems (P4-P6), performance has increased consistently,outofwhichP6gavebetterperformancethanP4and P5witharelativeperformanceof14.2%comparedtoDCASE2018 baseline system.It can be observed that the DNN score fusion of PLP, PLPCC and Log-Mel band energies feature in the proposed system (P7) has given significant improvement compared to the DCASE 2018 baseline system.Overall, 14.4% relative improve- mentisachievedwiththeproposedsystem. Usingproposedfeatures (P7)theBus,Park,StreettrafficandShoppingmallclassesarewell

Table3.Average(B,C)accuraciesforSubtaskB.

Accuracy	Baseline-2018[1]	P1	P2	P3	P4	P5	P6	P7
Average	45.6%(±3.6)	59.2%	58.1%	54.5%	59.3%	59.6%	59.8%	60.0%

Table4.AverageaccuraciesforDCASE2017dataseton4foldcrosssvalidation.

Accuracy	Baseline-2017[25]	P1	P2	P3	P4	P5	P6	P7
Average	74.8	80.6%	79.5%	75.9%	80.1%	80.4%	80.6%	80.8%

classifiedwhencomparedtootherclassesforsubtaskB.Further,we havealsoexperimentedwiththeotherproposedsystems(P1,P2,P3, P4, P5, P6, and P7).The average (B, C) performance was obtained with various proposed feature sets shown in Table 3.From the ta- ble, it can be observed that all the proposed systems have given an improvementintheaverage(B,C)accuracy.Therelativeimprove- mentsof13.6%,12.5%,8.9%,13.7%,14.0%,14.2%,and14.4%are
obtained for P1-P7 respectively as compared to the DCASE 2018 baseline system.

D. AnalysisofResultsonTUTAcousticScenes2017Ddataset

TheresultsonTUTAcousticScenes2017datasetforASCtaskare presented in Tables 4 and 5.
From the Table 4, individual PLP features perform better than PLPCC and Log-Mel features.For two features combination pro- posedsystems(P4-P6),performancehasincreasedconsistently,out of which P6 has given better performance than P4 and P5 with a relative performance of 5.8% compared to DCASE 2017 baseline system. Itcanbeobservedthatalltheproposedsystemshavegiven animprovementintheaverageaccuracy.Therelativeimprovements of5.8%,4.7%,1.1%,5.3%,5.6%,5.8%,and6.0%areobtainedfor P1-P7 respectively as compared to the DCASE 2017 baseline sys- tem.
TheresultsonTUTAcousticScenes2017datasetaregiven in Table 5 for the DCASE 2017 baseline and our proposed sys-tem, which is the DNN score fusion of PLP, PLPCC and Log-Mel bandenergiesfeatures(P7)andalsothecomparisonwithstate-of-art in [29].

Table5.ASCresults,averagedoverevaluationfoldsandcompari- son with DCASE 2017 Baseline system and state of the art [14].

AcousticScene	Baseline-2017[25](%)	Fusionwithoutaugmentation[14]	ProposedSystem (P7)(%)
Beach	75.3	70.9	51.3
Bus	71.8	82.1	87.2

Cafe/Restaurant	57.7	71.8	64.1
Car	97.1	89.0	100.0
CityCenter	90.7	85.6	94.9
ForestPath	79.5	97.3	47.4
GroceryStore	58.7	83.3	89.7
Home	68.6	76.0	89.7
Library	57.1	82.0	80.8
MetroStation	91.7	90.7	98.7
Office	99.7	95.1	97.4
Park	70.2	69.9	55.1
ResidentialArea	64.1	71.8	94.9
Train	58.0	71.8	69.2
Tram	81.7	84.6	91.0
Overallaccuracy	74.8	81.5	**80.8**

Itcanbeseenfromthetablethat,theproposedfeatures(P7),are well classified for Car, City center and Office classes compared to DCASE 2017 baseline [25] and [29].

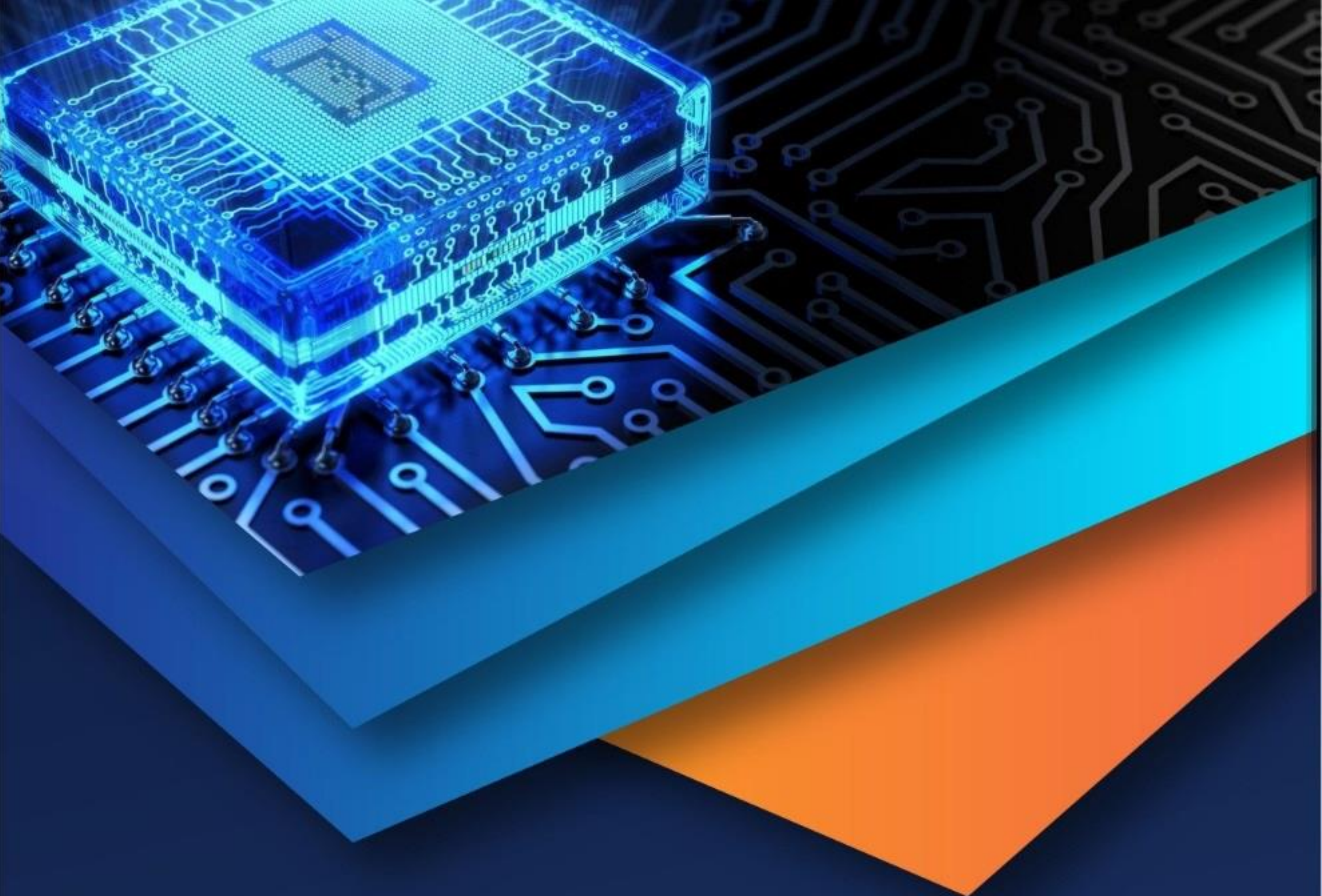
V. SUMMARY AND CONCLUSIONS

Inthispaper,aninvestigationofPLPandPLPCCfeatureswithDNN architecture has been applied to model the ASC. We experimented with TUT Acoustic Scenes 2018 Datasets of task1 including Sub- task A and B and TUT Acoustic Scenes 2017 dataset.The study demonstrated that the capability of individual feature sets and fu- sion of PLP, PLPCC and Log-Mel band energies at DNN score de- cisionlevel. IndividualPLPfeaturesyieldanimprovementof8.9% and 13.6% and PLPCC features result in an improvement of 8.6% and 12.5% for subtask A and subtask B of DCASE 2018 challenge. Significantimprovementsinaccuracyisachievedfor DNNdecision levelscoresobtainedfromPLP,PLPCCandLog-Melbandenergies. Improvementsof11.4%and14.4%wereachievedinsubtasksAand BrespectivelycomparedtotheDCASE2018ASCbaselinesystem. From DCASE TUT Acoustic Scenes 2017 dataset, individual PLP featuresyieldanimprovementof5.8%andPLPCCfeaturesresultin an improvement of 4.7% respectively.An improvement of 6.0% is achievedwiththefusionstudycomparedtotheDCASE2017base- linesystem.ThisshowsthatPLP,PLPCCandLog-Melbandener- giescarrycomplementaryacousticinformation. Futureworkwould be dedicated to the investigation of different combinations of fea- tures for ASC.

REFERENCES

[1] Mesaros, Annamaria and Heittola, Toni and Virtanen, Tuo- mas,“A multi-device dataset for urban acoustic scene clas- sification,”Submitted to DCASE2018 Workshop, 2018.
[2] Daniele Barchiesi, Dimitrios Giannoulis, Dan Stowell, and MarkD.Plumbley,“Acousticsceneslassification:Classifying environments from the sounds they produce,”The Journal of IEEE Signal Processing Magazine , vol. 32, pp. 16–34, 2015.
[3] SChu,SNarayanan,CJKuo,andMJMataric,“WhereamI? scene recognition for mobile robots using audio features,”in Proc.IEEEInternationalConferenceonMultimediaandExpo, 2006, pp. 885–888.
[4] J T Geiger, B Schuller, and G Rigoll,“Recognising acoustic sceneswithlarge-scaleaudiofeatureextractionandSVM,”in Proc. IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events, 2013.
[5] W Nogueira, G Roma, and P Herrera,“Sound scene identifi- cationbasedonMFCC,binauralfeaturesandasupportvector machineclassifier,”inProc.IEEEAASPChallengeonDetec- tion and Classification of Acoustic Scenes and Events, 2013.
[6] A. Rakotomamonjy and G. Gasso,“Histogram of gradientsof time-frequency representations for audio scene detection,” IEEETransactionsonAudio,Speech,andLanguageProcess- ing, vol. 23, no. 1, 2015.
[7] AxelPlinge,ReneGrzeszick,andGernotAFink,“ABag-of- featuresapproachtoacousticeventdetection,”inProc.IEEE InternationalConferenceonAcousticsSpeechandSignalPro- cessing (ICASSP), 2014, pp. 3704–3708.
[8] PKhunarsal,CLursinsap,andTRaicharoen,“Veryshorttime environmentalsoundclassificationbasedonspectrogrampat- ternmatching,”TheJournaloftheInformationSciences,vol. 243, pp. 57–74, 2013.

[9] JSalamonandJPBello,“Deepconvolutionalneuralnetworks and data augmentation for environmental sound classification,” vol. 24, no. 3, pp. 279–283, 2017.
[10] W Nogueira, G Roma, and P Herrera,“Sound scene identifi- cationbasedonMFCC,binauralfeaturesandasupportvector machineclassifier,”inProc.IEEEAASPChallengeonDetec- tion and Classification of Acoustic Scenes and Events, 2013.
[11] M Chum, A Habshush, A Rahman, and C Sang,“Scene clas- sification challenge using hidden Markov models and frame based classification,”in Proc. IEEE AASP Challenge on De- tectionandClassificationofAcousticScenesandEvents, 2013.
[12] HamidEghbal-Zadeh,BernhardLehner,MatthiasDorfer,and Gerhard Widmer,“CP-JKU submissions for DCASE-2016:a hybrid approach using binaural i-vectors and deep convolu- tionalneuralnetworks,”Tech.Rep.,DCASE2016Challenge, September 2016.
[13] Rakib Hyder, Shabnam Ghaffarzadegan, Zhe Feng, John H L Hansen,andTaufiqHasan,“Acousticsceneclassificationusing aCNN-Supervectorsystemtrainedwithauditoryandspectro- gram image features,”in Proc. INTERSPEECH, 2017.
[14] SeongkyuMun,SangwookPark,DavidHan,andHanseokKo, “Generativeadversarialnetworkbasedacousticscenetraining setaugmentationandselectionusingSVMhyper-plane,”Tech. Rep., DCASE2017 Challenge, September 2017.
[15] YPetetin,CLaroche,andAMayoue,“Deepneuralnetworks foraudioscenerecognition,”inProc.EuropianConferenceon Signal Processing (EUSIPCO), 2015, pp. 125–129.
[16] V Bisot, R Serizel, S Essid, and G Richard,“Acoustic scene classification with matrix factorization for unsupervised fea- ture learning,”in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2016, pp. 6445–6449.
[17] HongweiSongandJiqingHanandShiwenDeng,“Acompact and discriminative feature based on auditory summary statis- tics for acoustic scene classification,”in Proc. Interspeech, 2018, pp. 3294–3298.
[18] Manjunath Mulimani and Shashidhar G Koolagudi,“Robust acoustic event classification using bag-of-visual-words,”in Proc. Interspeech, 2018, pp. 3319–3322.
[19] Qian, Kun and Ren, Zhao and Pandit, Vedhas and Yang, Zi- jiangandZhang,ZixingandSchuller,Bjo¨rn,“Waveletsrevis- ited for the classification of acoustic scenes,”in Proc. Work- shop on Detection and Classification of Acoustic Scenes and Events (DCASE), November 2017, pp. 108–112.
[20] ShefaliWaldekarandGoutamSaha,“Wavelet-basedaudiofea- tures for acoustic scene classification,”Tech. Rep., DCASE Challenge, September 2018.
[21] ShefaliWaldekarandGoutamSaha,“Wavelettransformbased mel-scaledfeaturesforacousticsceneclassification,”inProc. Interspeech, 2018, pp. 3323–3327.
[22] Sakashita, Yuma and Aono, Masaki,“Acoustic scene classi- fication by ensemble of spectrograms based on adaptive tem- poral divisions,”Tech. Rep., DCASE Challenge, September 2018.
[23] Dorfer, Matthias and Lehner, Bernhard and Eghbal-zadeh, Hamid and Christop, Heindl and Fabian, Paischer and Ger- hard,Widmer,“Acousticsceneclassificationwithfullyconvo- lutional neural networks and I-vectors,”Tech. Rep., DCASE Challenge, September 2018.
[24] Zeinali, Hossein and Burget, Lukas and Cernocky, Honza, “Convolutional neural networks and x-vector embedding for dcase2018acousticsceneclassificationchallenge,”Tech.Rep., DCASE Challenge, September 2018.
[25] Annamaria Mesaros, Toni Heittola, and Tuomas Virtanen, “TUT database for acoustic scene classification and sound eventdetection,”inProc.24thConferenceonEuropeanSignal Processing (EUSIPCO), Budapest, Hungary, 2016.
[26] HynekHermansky,“Perceptuallinearpredictive(PLP)analy- sisofspeech,”TheJournaloftheAcousticalSocietyofAmer- ica, vol. 87, no. 4, pp. 1738–1752, 1990.
[27] Hacker Christian Hoenig Florian, Stemmer Georg and Brug- naraFabio, “RevisingPerceptualLinearPrediction(PLP),” in Proc. INTERSPEECH, 2005, pp. 2997–3000.
[28] J Schmidhuber,“Deep learning in neural networks: An overview,” arXiv, 2014.
[29] Diederik, P Kingma, and Jmmy Lei Ba,“ADAM: A Method for Stochastic Optimization,”in Proc. Conference on ICLR, 2015.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

TOGETHER WE REACH THE GOAL

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY