will be loaded into the workspace. Our proposed method is divided into four different parts: In the first part we have trained the audio signals and then silence has been removed from the audio signals. In the second part features were extracted from audios using I-Vector. In the third part split overlapping function is applied to evaluate the overlapped audio beats. In the fourth part we have evaluated depression using relationship matrix. Table I refers to nomenclature of metrics used in equations.

TABLE I. Nomenclature of metrics used in equations

Metric	Full Name
temp	A
Amp	Amplitude
F	Frequency
Spt	Split
Ovp	Overlapping
Ado	Audio
Len	Length
Inc	Increment
Abs	Absolute
Min	Minimum
Max	Maximum
Coef	Coefficient
Fft	Fast Fourier transform
Dist	Distance
Corr	Correlation
Tp	True Positive
tn	True Negative
Fp	False Positive
Fn	False Negative
Sum	∑
Ft	Filter
Frame Len	Fl
Enframe	B

i. Silence Removal

In this part we have removed silence from the audio signals using the following equations as given below. In equation 1 we calculated the sum of amplitude frequency of signal. In equation 2 and 3 we calculated the minimum amplitude frequency of signal to remove silence from the signal.

$$\mathrm{amp_f} = \sum \left(\mathrm{abs}\left(\mathrm{spt_{ovp}}(\mathrm{ado}([1 \; -0.9375], 1, \mathrm{x}), \mathrm{adoLen}, \mathrm{adoInc})\right), 2\right); (1)$$

$$\mathrm{amp_{f1}} = \min\left(\mathrm{amp_{f1}}, \frac{\max(\mathrm{amp_f})}{4}\right); (2)$$

$$\mathrm{amp_{f2}} = \min\left(\mathrm{amp_{f2}}, \frac{\max(\mathrm{amp_f})}{8}\right); (3)$$

Code for silence removal

```
if amp_fq(n) > amp_fq1
x1 = max(n-count-1,1);
elseif amp_fq(n) > amp_fq2 OR zcount(n) > zcount2
silent = silent+1;
if silent < peak_silent
silent = 0;
end
```

ii. Feature Extraction Using I-Vector

The In the second part we have extracted features from audio signals using I-Vector method using the following algorithm:

a. Initialize vector variable with Melvectorm function using equation 4.

$$\mathrm{vector} = \mathrm{melvectorm(audio)}; (4)$$

b. Calculate frequency to bit-ratio and store it in random variable l_r using equation 5.

$$l_r = \frac{\log\left(\frac{f_{zero} + fh}{f_{zero} + f1}\right)}{p + 1}; (5)$$

c. Convert l_r value to fast fourier transform bin numbers using equation 6, 7, 8, 9 and 10.

$$b1 = n * ((f_{zero} + f1) * \exp([0 \quad 1 \quad p \quad p + 1] * l_r) - f_{zero}); (6)$$

$$\mathrm{p_f} = \frac{\log\left(\frac{\mathrm{f_{zero}} + \frac{\mathrm{b2{:}\,b3}}{\mathrm{n}}}{\mathrm{f_{zero}} + \mathrm{f1}}\right)}{\mathrm{l_r}}; (7)$$

$$r = [ones(1, b2)fp \quad fp + 1 \quad P * ones(1, frq_{n2} - b3)]; (8)$$

$$c = [1{:}\,b3 + 1 \;\; b2 + 1{:}\,frq_{n2} + 1;]; (9)$$

$$\mathrm{v} = 2 * \left[0.5\backslash(1, \mathrm{b2} - 1)1 - \mathrm{p_f} + \mathrm{f_p} \;\; \mathrm{p_f} - \mathrm{f_p}(1, \mathrm{frq_{n2}} - \mathrm{b3} - 1\,)0.5\right]; (10)$$

d. Using equation 11 and 12 we calculated the value of melvectorm function.

$$\text{vector} = 1 - \frac{0.92}{1.08} * \cos\left(\text{v} * \frac{\text{pi}}{2}\right); \ (11)$$

$$vector = \frac{vector}{\max\left(vector(:)\right)}; \ (12)$$

e. In equation 13 and 14 we calculated the vector value of signal and store it in variable w.

$$w = 1 + 6 * \sin\left(pi * \frac{[1:12]}{12}\right); \ (13)$$

$$w = \frac{w}{\max(w)}; \ (14)$$

f. Store the result.

iii. Evaluate Overlapped Audio Beats

In this part we have evaluated overlapped audio beats using split overlapping function using the following algorithm:

a) Read the audio data.
b) Multiply the audio data with the hamming distance.
c) Apply the fast fourier transform function on the above data.
d) Calculate the distance vector and store it in distance vector variable.
e) Calculate correlation matrix.
f) Evaluate the overlapped audio beats using split overlapping function.
g) Store the result.

iv. Evaluate Depression Using Relationship Matrix

In this part we have evaluated depression using the Following equation:

$$relation(i,j) = sum\left(\left(t(i,:) - r(j,:)\right).^2\right); \ (15)$$

In this equation here t is the trained signal and r is the input signal. For given signal t, if the matched value of r signal is high then depression will be the depression value stored in the trained data set.

4.2 Estimation of Depression Level in Audios Using Fuzzy Membership Functions

The technique used in our analysis depicts audio data using fuzzy membership function. We have used fuzzy membership functions to make the method durable against various other sources in the audios. First of all the audios will be loaded into the workspace. Our proposed method is divided into five different parts: In the first part we have normalize the audio signals. In the second part we have evaluated amplitude by applying audio filtering. In the third part we have evaluated changes in amplitude audio. In the fourth part we have defined fuzzy rules. In the fifth part it will return membership values.

- Normalize Filter

$$x = \frac{x}{\max(abs(x))}; \ (16)$$

$$x1 = numel(x); \ (17)$$

$$\alpha 1 = \beta(x(1:end-1), fl, inc); \ (18)$$

$$\alpha 2 = \beta(x(2:end), fl, inc); \ (19)$$

$$signs = (\alpha 1.* \alpha 2), 0; \ (20)$$

$$diffs = (\alpha 1 - \alpha 2) > 0.02; \ (21)$$

$$zcr = sum(signs.* diffs, 2); \ (22)$$

- Evaluate amplitude by applying audio filtering

$$\text{amp} = \sum\left(\left(\text{abs}\left(\beta(\text{ft}([1 \ \ -0.9375], 1, \text{x}), \text{fl}, \text{inc})\right)\right), 2\right); \ (23)$$

$$\text{av}_{\text{zcr}} = \frac{\text{sum(zcr)}}{\text{len(zcr)}}; \ (24)$$

$$\text{av}_{\text{amp}} = \frac{\text{sum(amp)}}{\text{length(amp)}}; \ (25)$$

$$\max \text{amp} = \max\left(\text{abs}(\text{amp})\right); \ (26)$$

$$\min \text{amp} = \min\left(\text{abs}(\text{amp})\right); \ (27)$$

```
Case 2
if (Fvzcr(n) 0.5+NFvzcr(n) 0.3+NFvamp(n) 0.2)>0.8
 count = count + 1;silence=0;
 else
 silence = silence+1;
 if silence < maxsilence %
 count = count + 1;
 else
 status = 3;
 end
 end
 Case 3,
 break;
 end
end

count = count-silence;
x2 = x1 + count -1;
```

ii. Return Membership Values

$$\text{final}_f = \frac{\text{abs(x2+count)}}{\text{abs(x2*count)}}; \quad (28)$$

7. Visual Analysis

Figure1, Figure2 and Figure 3 are showing the input or actual signal as well as the test or matched signal and both the signals are very similar. After that we will calculate the value of accuracy, specificity, peak signal to noise ratio, f-measure and balanced classification rate for test signal. After calculating the values of these parameters we will estimate the depression level in each signal.

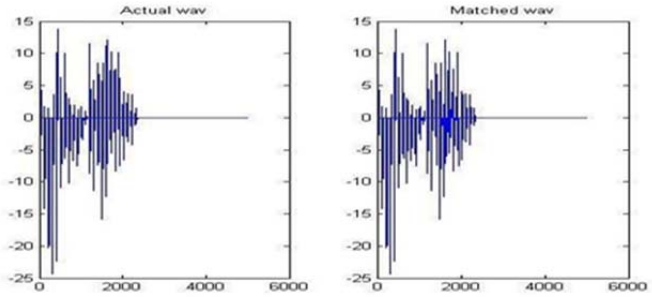

Fig.1 Diagram of actual waveform and matched waveform

8. Performance Evaluation

This section contains the comparison table and graphs of the existing and proposed techniques. Some well-known image performance evaluation parameters for audio signals have been selected to prove that the performance of the proposed algorithm is quite better than the existing method.

1. MAE – In statistics, the mean absolute error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The mean absolute error is given by:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| \quad (1)$$

The values of Mean absolute error are shown below in the comparison Table 5.1

Table 5.1: Mean absolute error comparison table

Input Audio Signals	Existing Result	Proposed Result
1	17.5924	0.3093
2	22.7627	1.5122
3	19.3833	1.5283
4	29.1466	8.1854
5	20.5230	0.6322
6	22.5241	1.2103
7	23.0225	1.2151
8	28.8334	7.3896
9	17.2772	1.5283
10	25.4083	4.4296

Figure 5.1 has shown the comparison table of the mean square error of different audio signals by Existing value in (Blue line) & proposed values in (Red lines). It is very clear from the graph that there is decrease in MAE value of signals with the use of proposed method over existing methods.

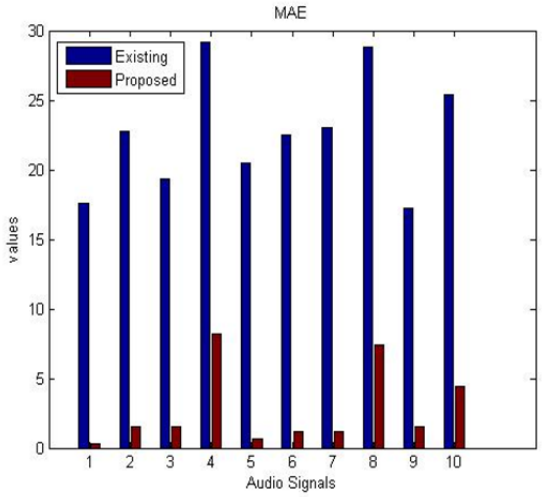

Figure 5.1: Mean Absolute error graph

This accuracy shows the accurate value of the input signal in case of the proposed algorithm as well as in existing algorithm.

Table 5.4: Accuracy comparison table

Input Audio Signals	Existing Result	Proposed Result
1	0.9569	0.9980
2	0.9533	0.9788
3	0.9565	0.9932
4	0.9467	0.9693
5	0.9557	0.9932
6	0.9441	0.9836
7	0.9499	0.9836
8	0.9293	0.9645
9	0.9617	0.9932
10	0.9305	0.9788

The values of accuracy are shown below in the comparison Table 5.4.

Figure 5.4 has shown the quantized analysis of the Accuracy. It is very clear from the plot that the value of Accuracy is getting maximized in every case with the use of proposed method over other methods.

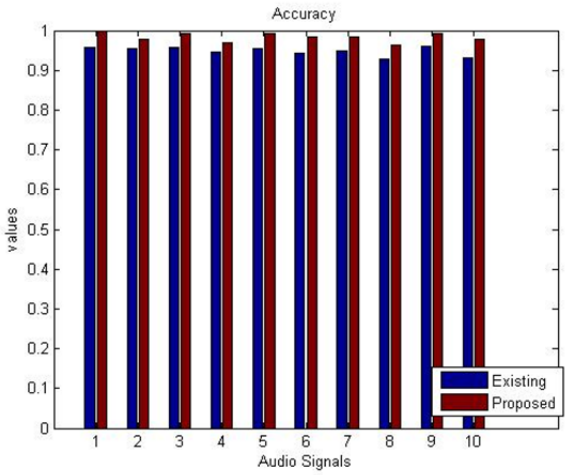

Figure 5.4: Accuracy graph

CONCLUSION

Depression has recently been attracting the attention of speech researchers, with audio/visual emotion challenge (AVEC) 2013 and 2014 organized to encourage researchers to develop approaches to accurately estimate speaker depression level. This dissertation has focused on speaker dependence of an I-Vector based depression level estimation system. I-Vector based depression level estimation system has better performance than existing techniques. In the existing depression level estimation techniques silence has not been removed from the input signal and fuzzy membership functions are also not used. In order to improve the accuracy of I-Vector based depression level estimation system we will design a fuzzy membership function and silence removal function in audio signals. The comparisons have clearly indicates that the proposed technique outperforms over the available techniques in terms of accuracy.

This work has not utilised any evolutionary optimization technique to match the trained depression audios with testing one so in near future we will utilize different evolutionary approaches to enhance the results further.

REFERENCES

[1] Paula Lopez-Otero, Laura Docio-Fernandez, Carmen Garcia-Mateo "Assessing speaker independence on a speech-based depression level estimation system" Pattern Recognition Letters (2015) Elsevier, pp. 1-8.

[2] P.Ghahremani, B.Baba Ali, D.Povey, K.Riedhammer, J.Trmal, S.Khudanpur, "A pitch extraction algorithm tuned for automatic speech recognition", in:Proceedings of ICASSP, 2014, pp.2494–2498.

[3] S.Alghowinem, R.Goecke, M.Wagner, J.Epps, G.Parker, M.Breakspear, "Characterising depressed speech for classification", in:Proceedings of Inter speech. ISCA, 2013, pp.2534–2538.

[4] N.Cummins, J.Joshi, A.Dhall, V.Sethu, R.Goecke, J.Epps, "Diagnosis of depression by behavioural signals: a multimodal approach", in: Proceedings of AVEC'13, 2013, pp.11–20.

[5] J.R.Williamson, T.F.Quatieri, B.S.Helfer, R.Horwitz, B.Yu,D. D.Mehta, "Vocal biomarkers of depression based on motor incoordination", in: Proceedings of AVEC' 13, New York, NY, USA, pp.41–48, 2013.

[6] M.Valstar, B.Schuller, K.Smith, F.Eyben, B.Jiang, S.Bilakhia, S.Schnieder, R.Cowie, M.Pantic, AVEC2013- "The continuous audio/visual emotion and depression recognition challenge", in: Proceedings of AVEC'13, 2013.

[7] L.J.Rodríguez-Fuentes, M.Penagarikano, A.Varona, M.Díez, G.Bordel, KALAKA 2: "a TV broadcast speech database for the recognition of Iberian languages in clean and noisy environments", in: Proceedings of LREC, pp.99–105, 2012.

[8] D.Povey, A.Ghoshal, G.Boulianne, L.Burget, O.Glembek, N.Goel, M.Hannemann, P.Motlicek, Y.Qian, P.Schwarz, J.Silovsky, G.Stemmer, K.Vesely,"The Kaldi speech recognition toolkit" , in: IEEE Workshop on Automatic Speech Recognitionand Understanding. IEEE Signal Processing Society, 2011.

[9] Niels Rosenquist, J., Fowler, J. & Christakis, N. "Social Network Determinants of Depression". Molecular Psychiatry 16 (3): 273–281, (2011).

[10] Kua, J. M. K., "Investigation of Spectral Centroid Magnitude and Frequency for Speaker Recognition." Proc. Odyssey: Speaker and Lang. Rec. Workshop, 2010, pp. 34 - 39.

[11] Low, L. S. A., N. C. Maddage, et al. "Mel frequency cepstral feature and Gaussian Mixtures for modeling clinical depression in adolescents." in Proc. IEEE Int. Conf. on Cognitive Informatics, 2009, pp. 346-350.

[12] Sethu, V., et al., ″Speaker dependency of spectral features and speech production cues for automatic emotion classification″, in Proc. IEEE ICASSP, 2009, pp. 4693-4696.

[13] McIntyre, G., R. Göcke, et al., "An approach for automatically measuring facial activity in depressed subjects", in Proc. Int. Conf. on Affective Computing and Intelligent Interaction and Workshops, 2009.

[14] E. Moore, et al. “Critical analysis of the impact of glottal features in the classification of clinical depression in speech,” IEEE Trans. Biomed. Eng., vol. 55, 2008, pp. 96-107.

[15] Yingthawornsuk, T., et al., "Direct Acoustic Feature Using Iterative EM Algorithm and Spectral Energy for Classifying Suicidal Speech." in Proc. Interspeech, 2007.

[16] Brierley, B.N. Medford., "Emotional memory for words: Separating content and context", Cognition & Emotion, 2007. 21(3): p. 495-521.

[17] Ozdas, A., “Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk”, IEEE Trans. Biomed. Eng., vol. 51, no. 9, 2004, pp. 1530- 1540.

[18] A.Rush, M.Trivedi, H.Ibrahim, T.Carmody, B.Arnow, D.Klein, J.Markowitz, P.Ninan, S.Kornstein, R.Manber, M.Thase, J.Kocsis, M.Keller, “The 16 item quick inventory of depressive symptomatology (QIDS) clinician rating (QIDS-C) and self-report (QIDS-SR):a psychometric evaluation in patients with chronic major depression”, Biol. Psychiatry 54, pp.573–583, 2003.

[19] France, D. J. “Acoustical properties of speech as indicators of depression and suicidal risk”, IEEE Trans. Biomed. Eng., vol. 47, no. 7, July 2000, pp. 829-837.

[20] Lustberg L, & Reynolds CF, “Depression and insomnia: questions of cause and effect. Sleep Medicine Reviews 4 (3): 253–262, (2000).