International Journal of Advanced Science and Technology
Vol. 29, No. 6s, (2020), pp. 1475-1482

Pushing the Boundaries of EEG-based Emotion Classification using Consumer-Grade Wearable Brain-Computer Interfacing Devices and Ensemble Classifiers

Jason Teo, Nazmi Sofian Suhaimi, James Mountstephens

Faculty of Computing & Informatics, Universiti Malaysia Sabah, Jalan UMS, 88450 Kota, Kinabalu, Sabah, Malaysia.

Abstract

Emotion classification using features derived from electroencephalography (EEG) is currently one of the major research areas in big data. Although this area of research is not new, the current challenge is now to move from medical-grade EEG acquisition devices to consumer-grade EEG devices. The overwhelmingly large majority of reported studies that have achieved high success rates in such research uses equipment that is beyond the reach of the everyday consumer. Subsequently, EEG-based emotion classification applications, though highly promising and worthwhile to research, largely remain as academic research and not as deployable solutions. In this study, we attempt to use consumer-grade EEG devices commonly referred to as wearable EEG devices that are very economical in cost but have a limited number of sensor electrodes as well as limited signal resolution. Hence, this greatly reduces the number and quality of available EEG signals that can be used as classification features. Additionally, we also attempt to classify into 4 distinct classes as opposed to the more common 2 or 3 class emotion classification task. Moreover, we also additionally attempt to conduct inter-subject classification rather than just intra-subject classification, which again the former is much more challenging than the latter. Using a test cohort of 31 users with stimuli presented via an immersive virtual reality environment, we present results that show that classification accuracies were able to be pushed to beyond 85% using ensemble classification methods in the form of Random Forest.

Keywords: *Emotion classification, EEG, Random Forest, Virtual Reality, Machine Learning*

1. INTRODUCTION

Emotion classification refers to the field of big data machine learning where brainwaves in the form of EEG signals are transformed into learning features that attempt to distinguish between a certain number of human emotions with reference to some particular model of emotions. One of the most commonly referenced models of emotion is known as Russell's Circumplex Model (Russell, 1980) where emotions can be grouped into four quadrants (classes) according to the two dimensions of valence and arousal (REF). The large majority of existing studies, in particular those that report high classification accuracy rates typically only conduct classification along one dimension (REF) only resulting in a two-class problem (e.g. high vs. low arousal or positive vs. negative valence) or a three-class problem (high, neutral, low arousal or positive, neutral, negative valence) (Alarcao & Fonseca, 2019). Here we attempt the significantly harder problem of distinctly classifying the emotions into four classes using both dimensions of arousal and valence.

Moreover, such training of the emotion classifier typically takes place for a particular user whereby the resulting emotion classifier is usable only by that particular user for previously unseen data (Alarcao & Fonseca, 2019). This method of emotion classification is known as intra-subject classification (REF). However, in this study, we attempt the more challenging emotion classification task of inter-subject classification whereby the trained emotion classifier can be used on users other than the user for which the training was conducted on. We further add on the challenge of conducting the classification using consumer-grade devices known as wearable EEG (Xu et al., 2017)) that cost

only USD199 with limited sensors and resolution as opposed to the large majority of successful studies that use medical-grade EEG BCI devices that typically cost upwards of USD10,000.
The objectives of this paper is threefold: (1) to demonstrate the usability of consumer-grade wearable EEG headsets as cost-effective and affordable deployment devices for affective computing applications; (2) to demonstrate the applicability of such devices for performing four-class emotion classification; and (3) to demonstrate the applicability of such devices for performing inter-subject emotion classification.

This paper is presented as follows: an overview of the research is presented in section 1, followed by section 2 which presents the background literature that motivates this study, section 3 presents the methods adopted in conducting this study, section 4 presents the results and analysis of the experiments conducted, and section concludes this study with the main findings and some future work.

2. BACKGROUND

Emotion classification studies have demonstrated the applicability of EEG-based sensing as a non-invasive mode of affective computing, with highly promising successful reports from diverse application domains such as music, images, and videos. However, the large majority of these studies use so-called medical-grade BCI devices which are typically only available in hospitals and research laboratories, thereby severely limiting the possibility of deploying emotion classification solutions in the real world. There are two main entry barriers to the deployment of such devices: (1) the high cost involved in acquiring the EEG acquisition equipment, (2) the prolonged set-up time and practical inconvenience involved in actually prepping the end-user for EEG acquisition. The costs of some of the BCI equipment used in such reported studies typically cost anywhere from USD1000 to upwards of USD25000 (Farnsworth, 2019). Hence, the average consumer would almost certainly not be willing to add on another interfacing equipment just to enable affective computing in their applications on top of the costs of the application itself.

Secondly, the large majority of such medical-grade BCI equipment utilizes a wet-electrode approach to signal acquisition, where a wet gel is required to be applied to the electrode before attaching to the user's scalp. This is usually a process which requires a significant amount of time to set up and additionally, later on for the user to clean their hair and scalp of the wet electrode gel after the usage of the BCI device. Clearly, this time-inefficient and inconvenient approach does not allow for a practical application of EEG-based sensing to the average home consumer in terms of practicable affective computing solutions. Consequently, a more cost-effective and practical solution to deploying EEG-based affective computing BCI hardware is in order to make these emotion classification solutions desirable for consumption by the general public. In this respect, consumer-grade EEG BCI devices that cost than USD200 and uses dry electrode technology are highly desirable and they are now commonly available in the form of wearable EEG headset or headbands (REF). However, their applications are largely reserved for meditation purposes (Brinson, 2017).

Emotion classification research has also largely been skewed towards two-class classification within the dimensions of arousal or valence exclusively. Only more recent studies have begun to report results beyond the simple two-class classification problem. Existing literature has also clearly indicated that the more successful results come from training and testing the emotions classified from within the same participant, which is known as intra-subject or within-subject emotion classification. When training is done on a participant and later tested on another participant, known as inter-subject or cross-subject emotion classification, the results are almost always inferior to the former, and the differences are significant. From the very few studies that have conducted both types of emotion classification and compared between both approaches within the same study, inter-subject classification accuracy rates were always lower than intra-subject classification accuracy rates where the difference was typically between 15-25% (Alarcao, 2019). Although inter-subject classification is clearly more challenging than intra-subject emotion classification, again for real-world deployment to everyday consumers of affective computing application, EEG-based emotion classification solutions need to be able to achieve reasonably high accuracy rates in inter-subject classification since no new retrainings of existing classification models need to be carried prior to its use.

The large majority of emotion classification studies focus on individual classifiers such as support vector machines, k-nearest neighbour, neural networks and deep learning approaches. There have been significantly less emotion classification studies that have attempted to use ensemble classifiers such as Random Forest. Some studies have used Random Forest to successfully classify emotion from speech (Badshah et al., 2016), music lyrics (Rachman et al., 2018), facial recognition (Jayalekshmi & Mathew, 2017), and tweets (Vora et al., 2017). Essentially Random Forest is a collection of individual decision trees that perform together as an ensemble (Liaw & Wiener, 2002). Its output is the majority decision obtained from the collection of individual decision trees.

3. METHODS

This section presents the methods adopted to conduct the investigations as explained in Section 1.

3.1 Emotional Stimuli

As the objective of this study is to classify the emotional response into four distinct quadrants according to Russell's Circumplex Model of Emotions, YouTube 360 videos were carefully selected to evoke participants' stimuli from each of the four quadrants according to the presentation protocol illustrated in Fig. 1:

(i) Quadrant 1 (Top right): High Arousal, Positive Valence (HA/PV) -> HAPPY
(ii) Quadrant 2 (Top left): High Arousal, Negative Valence (HA/NV) -> UPSET
(iii) Quadrant 3 (Bottom left): Low Arousal, Negative Valence (LA/NV) -> BORED
(iv) Quadrant 4 (Bottom right): Low Arousal, Positive Valence (LA/PV) -> RELAXED

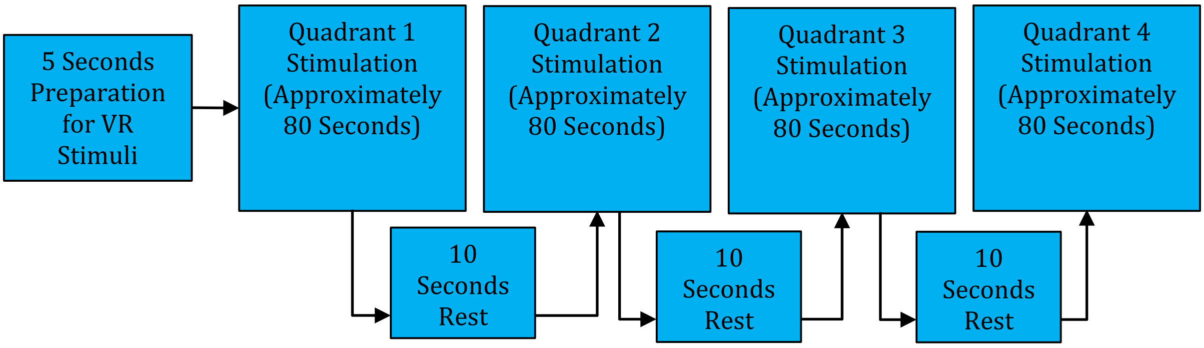

Fig. 1. Presentation of VR stimuli to participants

3.2 Brain-Computer Interfacing

Using EEG to capture the participants' brainwaves as they are stimulated by and reacting to the VR content emotionally, a BCI device called Muse (Krigolson, 2017) was chosen for this purpose as shown in Fig. 2. This version utilizes the Bluetooth Low Energy (BLE) 4.0 communication protocol and is required for pairing the connection to a smartphone for transmission and acquisition of the EEG signals from Muse wirelessly.

Fig. 2. Muse EEG Headset

The Muse headset is worn around the front of the head similar to a conventional headband. It has one reference electrode at the Fpz and four recording electrodes positioned at AF7, AF8, TP9, and TP10 according to the international 10-20 positioning system (Herwig et al., 2003). An accompanying app provides the functionality of pre-processing the raw EEG data such as using the notch filter at 60Hz for removal of artifacts as well as obtaining the absolute band powers for the five frequency bands commonly used in the literature for the study of human brain activity (Buzsaki, 2006), namely the:

- (i) Delta (0Hz to 4Hz);
- (ii) Theta (4Hz to 8Hz);
- (iii) Alpha (8Hz to 13Hz);
- (iv) Beta (13Hz to 30Hz); and
- (v) Gamma (30Hz to 44Hz) bands.

The absolute band power for a given frequency range is calculated as the logarithm of the sum of the Power Spectral Density (PSD) of the EEG data over a particular frequency range of the specified band. They are provided for each of the four to six channels/electrode sites on Muse. A Fast Fourier Transform (FFT), which is an algorithm which computes the Discrete Fourier Transform (DFT) of a given sequence, is used to compute the PSD of the EEG data with a window size of 256 samples with an overlap of 90% which is sampled at 220Hz providing a frequency resolution of $220/256 \approx 0.86$Hz/bin. Muse also provides the functionality of recording the raw EEG signals and well as supplementary sensor readings from its accelerometer and gyroscope. In this study, only the raw and discrete-time FFT was used in the classification tasks.

3.3 Machine Learning Environment and Classification Algorithms

Deep learning (LeCun et al., 2015), Naïve Bayes (Rish, 2001), and Random Forest (Liaw & Wiener, 2002) classification algorithms were used for the emotion recognition learning task. The R programming language environment is used to perform the machine learning tasks required in this investigation. It was also used for the evaluation of the experimental results to facilitate the analysis of the data collected from the participants. The main application programming interface (API) library used in R for this machine learning investigation was the "h2o" library, which implements Deep Learning (DL) classifiers and where the R h2o library represents interfacing code with the original Java codes in which the h2o deep learning library was written. Using the "caret" machine learning library in R, the Naïve Bayes (NB) and Random Forest (RF) approaches were also investigated for classification. These classifiers were run using the default settings found in the "caret" library. For Random Forest, the "ntree" parameter was kept at the default setting of 500 and the "mtry" parameter is automatically calculated by the algorithm based on the number of input features.

3.4 Deep Learning Parameters

In order to determine the suitable parameters to use for the deep neural network architectures to be used in the actual experimental runs, a preliminary test was conducted. The deep neural networks used for classification in this preliminary test yielded architectures that performed best when there were set to two hidden layers with 200 hidden nodes within each layer. As mentioned before, 10-fold cross-validation was used for each of the deep neural networks tested and these were individually run each time using for 10 epochs in each experiment. The uniform adaptive method was used to initialize the weights of the deep neural networks while the cross-entropy error function was used to assess the deep neural networks during training and testing phases. The Rectified linear unit (ReLU) transfer function was implemented as the deep neural network's activation function and its dropout ratio parameter was set at 50%. An adaptive learning rate was also implemented for all the hidden layer nodes while the output layer was implemented with a softmax transfer function.

4. RESULTS AND DISCUSSION

This section presents the results obtained from the experiments conducted. All results are presented for inter-subject classification for four-class emotion classification. The results are presented as follows: (1) 10-fold cross-validation average classification results by channels and sensors; (2) graphical comparison of classification accuracy between the different frequency bands used as input features to the classifiers; and (3) distribution of inertial sensor data against emotional stimuli quadrant.

4.1 Channel and Sensor Analysis

Table 1. 10-fold Cross-Validation Average Classification Accuracies by Channel and Sensor.

Channels/Sensors Used	**Accuracy**		
	DL	NB	RF
Raw EEG Frontal	25.14%	27.62%	26.93%
Raw EEG Temporal-Parietal	26.03%	28.61%	26.93%
Raw Frontal + Temporal-Parietal	26.77%	29.85%	29.57%
Full FFT Spectrum Frontal	36.06%	40.65%	73.04%
Full FFT Spectrum Temporal-Parietal	35.66%	38.51%	86.18%
Full FFT Spectrum Frontal + Temporal-Parietal	***42.55%***	*44.33%*	***86.20%***
Raw EEG + Full FFT Spectrum Frontal	33.06%	33.90%	72.01%
Raw EEG + Full FFT Spectrum Temporal-Parietal	34.29%	38.84%	73.40%
Raw EEG + Full FFT Spectrum Frontal + Temporal-Parietal	39.09%	***44.49%***	84.26%
Accelerometer	39.64%	41.09%	43.36%
Gyroscope	29.05%	27.85%	29.53%
Accelerometer + Gyroscope	40.32%	41.08%	47.05%

Table 4.1 presents the average accuracy from ten-fold cross-validation of the four-class emotion classification task for an inter-subject classification approach. The best result of 86.20% was obtained using Random Forest with features comprising the full FFT spectrum of all five frequency bands that were obtained from both the frontal and temporal-parietal lobe sensors. This was followed closely by the results obtained also from Random Forest at 86.16% when using the full FFT spectrum of all five frequency bands that were obtained but only from the temporal-parietal lobe sensors. The next best results came from the addition of the raw EEG to the first results of the full FFT spectrum of all five frequency bands from both frontal and temporal-parietal lobe sensors at 84.26%. These are highly encouraging results that clearly prove that consumer-grade wearable EEG devices such as Muse with only four electrode sensors and with lower signal resolution compared to medical-grade EEG devices are able to produce very promising emotion classification accuracies of more than 80% using a variety of classification features for the highly challenging task of four-class, intra-subjection emotion classification.

4.2 Frequency Band Analysis

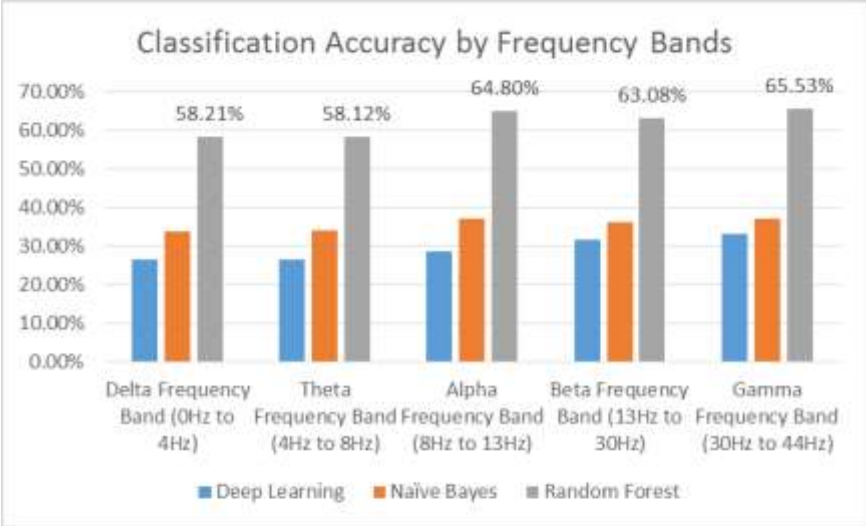

Fig. 3. Comparison of Classification Accuracy by EEG Frequency Bands

Next, the analysis from using individual frequency bands, namely the delta, theta, alpha, beta, and gamma frequency bands, obtained from the Discrete-time Fourier Transform is presented in Fig. 3. Clearly, again the Random Forest classifier provided the best classification across all frequency bands. The best results were obtained from the gamma band at 65.53%, followed by the alpha band at 64.80% and then the beta band at 63.08%. The classification accuracy obtained from the delta and theta bands was the lowest at just above 58%. This finding is in line with a number of recent studies that have indicated promising emotion classification results obtained from the higher frequency bands, in particular from the gamma band.

4.3 Inertial Sensing Analysis

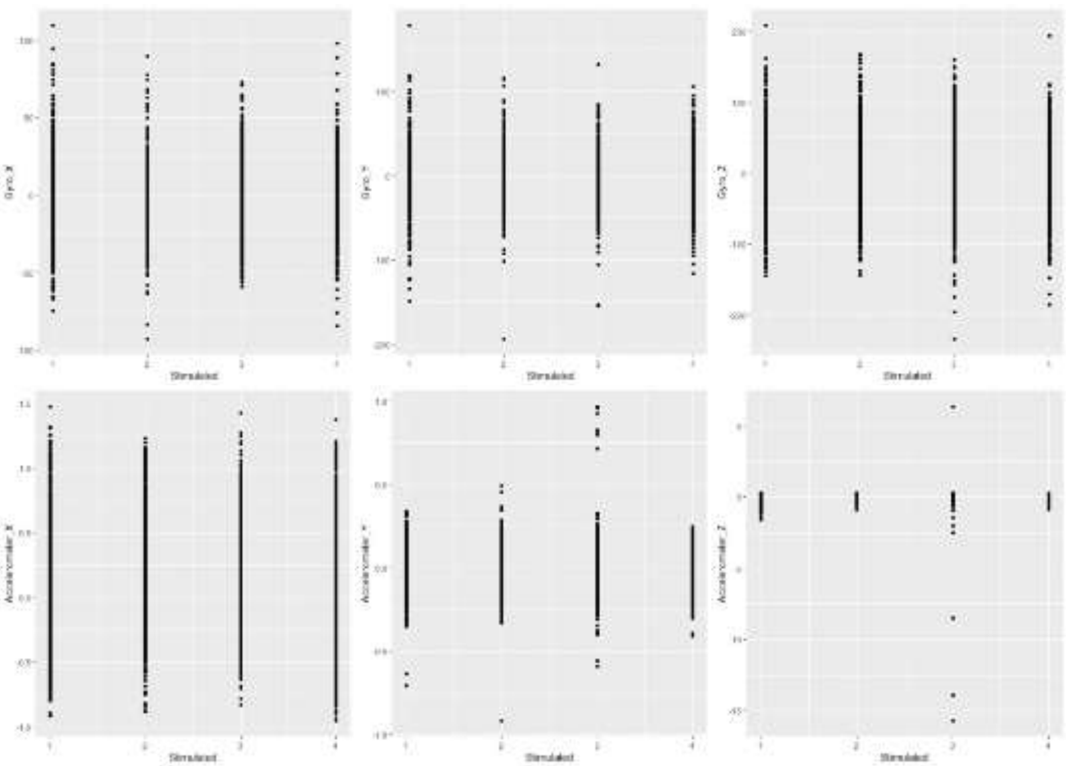
Fig. 4. Distribution of inertial sensor data against stimuli quadrant

ISSN: 2005-4238 IJAST

The raw coordinate in the x, y, and z planes are plotted by emotional stimuli quadrants in Fig. 4 above for both inertial sensing modalities acquired from the accelerometer and gyroscope sensors on the Muse device. Although the prediction accuracies presented earlier for these two sensor modalities do not provide as high a classification rate as the EEG sensors, there still appears to be some variation in head movements depending on the stimuli presented. It was interesting to observe that the gyroscope readings in the x-axis for the “bored” quadrant (labelled as quadrant no. 3) were had a less variable reading compared to the other quadrants indicating that there was less movement of the head about the x-axis.

Of more interest would be the accelerometer sensor readings which clearly show a clear pattern in terms of the magnitude of movements that were uniformly observed across all stimuli quadrants. Movements on the x-axis were the most significant compared to the y-axis, which comes next in terms of magnitude, followed lastly by the z-axis which had very minimal readings except for a few data points that were captured during the third quadrant stimuli presentations. This indicates that movements along the x-axis are noticeably large and provides an added modality for emotion detection applications.

5. CONCLUSION AND FUTURE WORK

The main objective of this study was to classify emotions triggered by virtual stimuli into four separate classes according to the arousal and valence dimensions of emotions using an inter-subject training and testing approach paired with a consumer-grade wearable EEG BCI device. Using an ensemble classifier approach in the form of Random Forest, the highest classification accuracy achieved for this set up was 86.20% with input features coming from all five frequency bands of both the frontal and parietal-temporal sensor electrodes. The results obtained using Random Forest outperformed Naïve Bayes and Deep Learning. For future work, eye-tracking technology is proposed to be used to augment the current EEG and inertial-sensing-based approach since in a virtual environment, eye gaze and stimuli localization could play an as yet unknown yet important role in VR emotion classification.

ACKNOWLEDGMENTS

This work was supported in part by the research grant ref: IF0318M1003 received from the Ministry of Energy, Science, Technology, Environment & Climate Change (MESTECC), Malaysia.

REFERENCES

[1] Alarcao, S. M., & Fonseca, M. J. (2019). Emotions recognition using EEG signals: A survey. *IEEE Transactions on Affective Computing*, *10(3)*, 374–393.

[2] Badshah, A. M., Ahmad, J., Lee, M. Y., & Baik, S. W. (2016). Divide-and-conquer based ensemble to spot emotions in speech using MFCC and random forest. arXiv preprint arXiv:1610.01382.

[3] Brinson, S. (2017). https://www.diygenius.com/hacking-your-brain-waves/

[4] Buzsaki, G. (2006). Rhythms of the Brain. Oxford University Press.

[5] Farnsworth, B. (2019). https://imotions.com/blog/eeg-headset-prices/

[6] Herwig, U., Satrapi, P., & Schönfeldt-Lecuona, C. (2003). Using the international 10-20 EEG system for positioning of transcranial magnetic stimulation. Brain topography, 16(2), 95-99.

[7] Hussain, A., Hashim, N. L., Nordin, N., & Tahir, H. M. (2013). A metric-based evaluation model for applications on mobile phones. *Journal of Information and Communication Technology*, *12*(1). 55-71

[8] Jayalekshmi, J., & Mathew, T. (2017). Facial expression recognition and emotion classification system for sentiment analysis. In 2017 International Conference on Networks & Advances in Computational Technologies (NetACT) (pp. 1-8). IEEE.

[9] Krigolson, O. E., Williams, C. C., Norton, A., Hassall, C. D., & Colino, F. L. (2017). Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. Frontiers in neuroscience, 11, 109.
[10] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18-22.
[11] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436.
[12] Rachman, F. H., Sarno, R., & Fatichah, C. (2018). Music emotion classification based on lyrics-audio using corpus based emotion. International Journal of Electrical and Computer Engineering, 8(3), 1720.
[13] Rish, I. (2001). An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41-46).
[14] Russell, J. A. (1980). A circumplex model of affect. Journal of personality and social psychology, 39(6), 1161.
[15] Vora, P., Khara, M., & Kelkar, K. (2017). Classification of Tweets based on Emotions using Word Embedding and Random Forest Classifiers. International Journal of Computer Applications, 178(3), 1-7.
[16] Xu, J., Mitra, S., Van Hoof, C., Yazicioglu, R. F., & Makinwa, K. A. (2017). Active electrodes for wearable EEG acquisition: Review and electronics design methodology. IEEE reviews in biomedical engineering, 10, 187-198.

ISSN: 2005-4238 IJAST