Modeling and Detecting Student Attention and Interest Level using Wearable Computers

Ziwei Zhu*, Sebastian Ober§, Roozbeh Jafari*
*Embedded Systems and Signal Processing Lab, Texas A&M University, College Station, TX 77840
§Department of Engineering and Gonville & Caius College, University of Cambridge, Cambridge, UK

***Abstract*— The cognitive states of students in a lecture can give good indications of student concentration and learning, and therefore, modeling them would have a positive impact on their quality of education by enabling the intervention of instructors. In a traditional class, the instructor would assess the students' level of attention. However, the assessment may not be accurate for a variety of reasons. Additionally, this creates a burden for the instructors. Wearable sensors and signal processing techniques could provide opportunities to assist teachers with this assessment. In this paper, we propose a methodology to model students' cognitive states by leveraging hand motion and heart activity captured with smart watches. Following the application of a sequence of signal processing techniques to the raw data, we generate features, which describe characteristics of the hand motion and heart activity in a group of students. The most prominent features are selected for machine learning algorithms. By applying cross validation, the results of experiments on 30 students in two lectures offer accuracies of 98.99% and 95.78% for predictions of 'interest level' and 'perception of difficulty' on the topics covered during the lectures.**

***Keywords*— cognitive states; inertial sensors; PPG; feature selection; classification**

I. Introduction

Researchers have suggested that identifying the quality of learning by monitoring the cognitive states of students is of value [1]. Teachers can adjust their teaching plans and methods in response to student cognitive feedback to improve the students' learning. For example, if students show little interest or feel the cognitive workload is too heavy, learning may be negatively impacted. It is therefore crucial for educators to understand the students' cognitive states during lectures. Traditionally, teachers rely on their experience, which may not be reliable in certain cases. Our work in this paper aims to solve this problem in a smart and automatic way by leveraging wearable sensors.

With the rapid development of sensors and signal processing technologies, sensor-assisted applications are increasingly being used in various scenarios, including educational applications. Examples of such technologies are context ubiquitous learning [2-8], augmented reality [9-13], and affective and cognitive states monitoring. Our work in this paper belongs to the last category. Some related studies have preceded our work. Conati measured student emotional states by a dynamic decision network with sensors measuring heart rate, skin conductance and eyebrow position [14]. Meredith combined a facial expression camera, a posture analysis seat sensor, an EEG sensor and an eye-tracker to model three affective states and three cognitive states [15]. Similar studies used the same sensors as the two projects above to recognize student affect and concentration [16, 17]. Mirko used cameras to mimic large-scale gaze tracking in order to model students' attention states [18].

Most previous work focused on the behavior of individuals. By contrast, our work in this paper proposes a different method to study classroom context. Second generation Moto 360 smartwatches, which contain 3-axis accelerometers, gyroscopes, and photoplethysmographic (PPG) sensors are attached to students' wrists with the aim of offering group level analyses. Fig. 1 shows an example of the study's scenario. Our work has the following contributions: 1) a high-performance hand-writing detection method is proposed, 2) multiple signal processing techniques applied to the sensors' data to generate high-level features which can describe the state of the group are designed and developed, and 3) our experiments achieved 98.99% and 95.78% accuracy (tested by the leave-three-out validation method) for the students' 'interest level' and 'perception of difficulty' cognitive states using 4 classification algorithms.

Figure 1. Example of Proposed System

The remainder of this paper is organized as follows: The proposed approach, including hand-writing detection, feature extraction and model-building are explained in Section II. In Section III, the experimental setup is introduced and the experimental results are described. Finally, the conclusion is provided in Section IV.

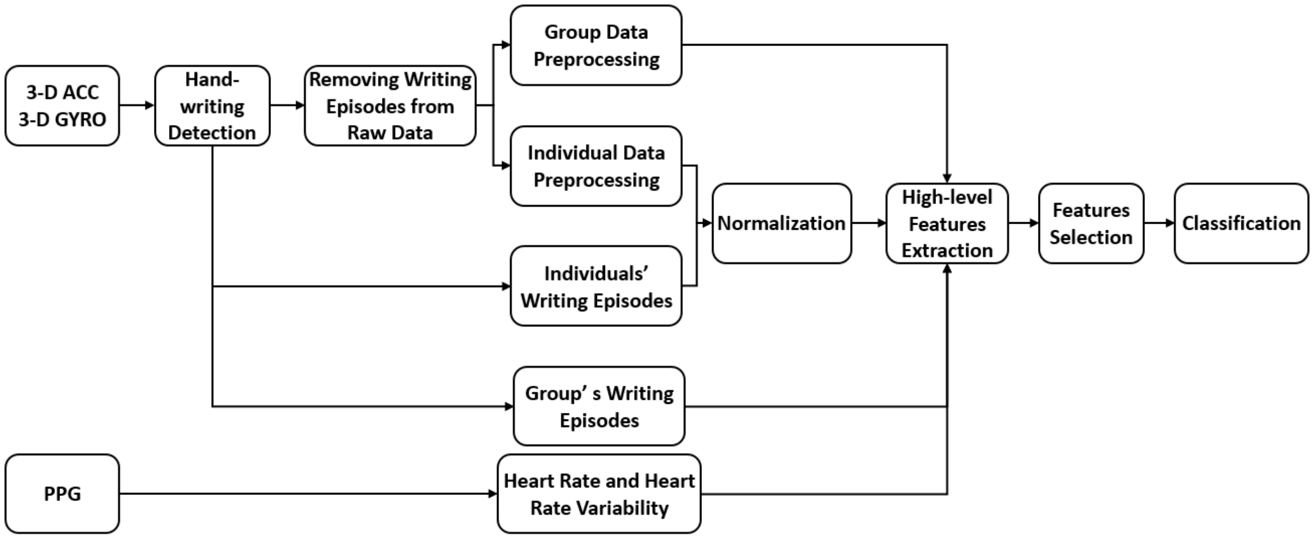

Figure 2. Diagram of Proposed Approach

II. Proposed Approach

Fig. 2 shows the diagram of the proposed approach to model students' interest level and perception of difficulty for various topics in a lecture. The entire process can be divided into five steps: 1) implementing hand-writing detection, 2) generating preprocessed series for individuals and the group from the raw data (excluding the writing episodes), 3) extracting several high-level features from the preprocessed series and results of writing detection, 4) applying feature selection techniques to the original feature set, 5) building models from the selected features by using four classification algorithms.

A. *Writing detection*

Through visual observations, we hypothesized that the action of writing is strongly correlated with students' cognitive states. It is reasonable that if students are interested in or feel that a slide of the lecture is hard to understand, they will take notes for that slide. Thus, hand-writing detection is the first step. It is also important to note that in the specific lectures that we observed, the instructor discourages the use of laptops and computers and hence the action of typing was not of interest.

Fig. 3 shows raw data from the 3-axis accelerometer and gyroscope during writing. Fig. 4 shows raw data during non-writing episodes for the purpose of comparison. The sensor in the smart watch is placed on the wrist, with the z-axis perpendicular to the wrist, facing outwards, the x-axis along the arm and the y-axis perpendicular to both the x- and z- axes. We observed that the accelerometer data remains at relatively constant levels because the wrist keeps the same pose during writing. Additionally, the acceleration due to gravity is greater than that due to the writing motion. As for the gyroscope, writing will affect certain axes more than others. In fact, the x-axis will give the largest variation because the wrist will rotate about the gyroscope's x-axis when writing. As a result of this analysis, we selected the mean values for the 3 accelerometer axes and the Mean Cross Rate values for the 3 gyroscope axes as the features. However, before feature extraction, segmentation would be required. Since writing is an activity that maintains the same motion from the beginning to the end, we chose a sliding window with the window size of 3 seconds. Then, we used the One-Class Support Vector Machine algorithm [19] with a non-linear kernel to build the classifier.

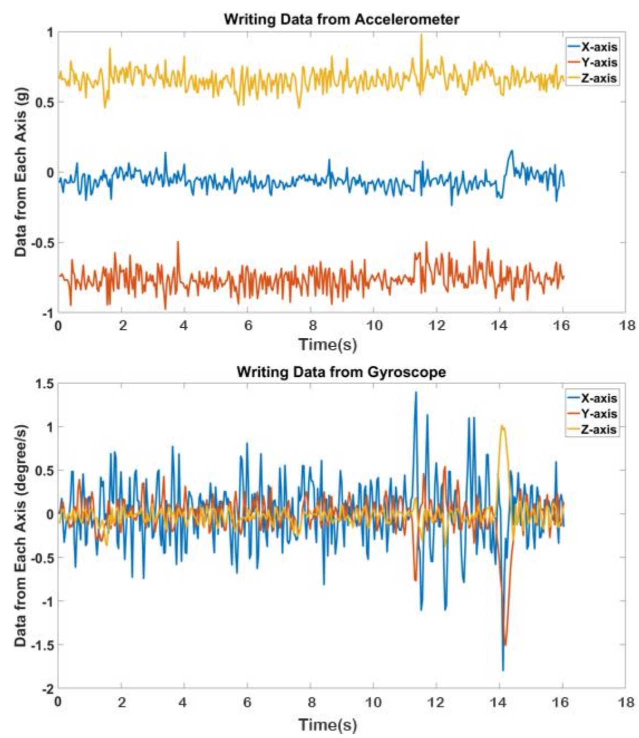

Figure 3. Data from Accelerometer and Gyroscope when Writing

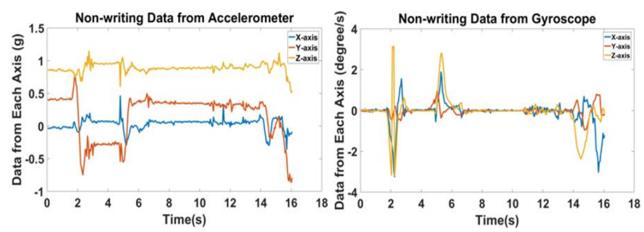

Figure 4. Data from Accelerometer and Gyroscope when Non-writing

TABLE I. RESULT OF HAND-WRITING DETECTION

Classification Report		
Precision	***Recall***	***F1-score***
0.9899	0.9800	0.9849

To assess the accuracy of our hand-writing episode detection, we collected data with motion sensors sampling at 25 Hz, and asked a user to perform 40 minutes of writing and 10 minutes of non-writing. 30 minutes of writing data was used for training, and 10 minutes of writing data and 10 minutes of non-writing data were used for the testing set. The experimental results of hand-writing detection are shown in Table I, which are suitable for the purpose of our signal processing development.

After applying the constructed classifiers to the raw data acquired from the smart watches in lectures, we identified the episodes of writing and non-writing for every student and time-stamped them. We combined individual student series to create group-level writing episodes. Fig. 5 shows an example where at any time, the number of students who are writing can be observed.

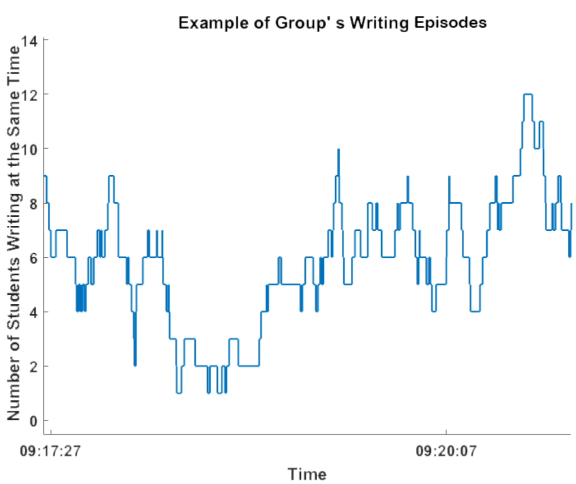

Figure 5. Example of Group' s Writing Episodes

B. Preprocessing the raw data without writing episodes.

Because the data associated with writing episodes may influence the extraction of other useful information about the group, we removed the writing episodes from the raw data. Since the raw data time-series are too basic and noisy to provide useful information, we applied certain preprocessing steps on the accelerometer and gyroscope data to prepare it for the next step, which involves feature extraction. We generated five time-series for every individual: an amplitude series, an entropy series, an estimated magnitude series, a heart rate series and a heart rate variability series.

The amplitude series gives the magnitude of the acceleration and angular velocity vectors, which are derived from the accelerometer and gyroscope respectively. Estimated magnitude and entropy are calculated from the amplitude series using a sliding window. In order to compute the estimated magnitude, we first divided the entire range of amplitude values into 50 intervals. We subsequently took a sliding window of 5 seconds in length to calculate a time array of means of the amplitude series. Finally, since each value of the mean array will belong to one of the 50 amplitude intervals, we approximated the values of the mean array with the medians of their corresponding intervals. The estimated magnitude can describe how large a student's motion is. The entropy indicates how large the variation of a student's motion is. We hypothesize that when a student is focusing on the lecture, the hand motions will decrease, and therefore the estimated magnitude and the entropy will decrease. In addition, we constructed similar time-series for the group: group entropy and group estimated magnitude. For these two series, we first merged all the individual amplitudes into one array and sorted it by time stamps. Then estimated magnitude and entropy are calculated from it using a sliding window of 5 seconds. By this method, we can get group level information about student hand motions.

Besides the hand motion, we attempted to study physiological observations acquired with smart watches from students. We collected PPG data at a frequency of 12.5 Hz. The PPG data was inherently noisy due to the presence of motion artifacts [20]. We applied a bandpass filter. We calculated the heart rate approximately every second by considering the peak-to-peak time intervals in the PPG signal. To reduce the impact of motion artifacts, a time-varying confidence measure at time t, $c(t)$, is introduced in Eq. (1), where $a(t)$ is the magnitude of the total acceleration at time t.

$$c(t) = \exp(-(a(t) - 9.8)^2) \qquad (1)$$

Therefore, the confidence associated with a heart rate value is unity when there is no net acceleration, and approaches zero as the net acceleration increases. However, the instantaneous confidence is not suitable as there may be instants when there is no net acceleration during a high-movement period. Therefore, for each point in the heart rate series, we took the lowest confidence near that timestamp. We used the confidence measures for mixing the individual heart rates into the group heart rate series, by using them to create a weighted average of the individual heart rates. We also averaged the measurements to create a group confidence series.

TABLE II. SERIES FROM PREPROCESSING

Series Name	Formula
Amplitude	$\mathrm{Amp} = \sqrt{x^2 + y^2 + z^2}$
Estimated Magnitude	$\mathrm{EM} = \mathrm{approximate}(\mathrm{mean}(\mathrm{window}(\mathrm{Amp})))$
Entropy Series	$\mathrm{En} = -\sum \mathrm{P} \times \log \mathrm{P}(\mathrm{mean}(\mathrm{window}(\mathrm{Amp})))$
Group Estimated Magnitude	$\mathrm{GEM} = \mathrm{approximate}(\mathrm{mean}(\mathrm{window}(\mathrm{mix}(\mathrm{Amp}))))$
Group Entropy	$\mathrm{GEn} = -\sum \mathrm{P} \times \log \mathrm{P}(\mathrm{mean}(\mathrm{window}(\mathrm{mix}(\mathrm{Amp}))))$
Heart Rate	$\mathrm{HR} = \mathrm{frequency} \div 60 \times \mathrm{time\_between\_peaks}$
Heart Rate Variability	SDANN and SDNNIDX with four different intervals

In addition to heart rate, we measured the heart rate variability (HRV) for each student. We used Modified SDANN and SDNNIDX to give series of heart rate variability, where SDANN is the standard deviation of means of NN intervals calculated over short sub-intervals, and SDNNIDX is the mean of standard deviations of NN intervals calculated over short sub-intervals [21]. We calculated both SDANN and SDNNIDX for every 30 seconds, minute, two minutes, and five minutes with five sub-intervals for each. Table II lists all the series and their formulas.

C. High-level features from preprocessed series and writing detection

Determining student cognitive states requires high-level features from the preprocessed series and the hand-writing episodes. Before feature extraction, normalization is required for both the preprocessed series and hand-writing detection results because individuals' behaviors in class vary. For example, some students take notes even if they think the current content is trivial, but they may take more notes when it is interesting (or difficult). On the other hand, other students may only take notes when they are interested in the content. Without normalization, the data from students who prefer to take more notes will override that from other students when combining all the students' writing episodes together. This will result in a loss of information about the group. The same applies to estimated magnitude and entropy. The normalization is shown in Eq. (2), where *Series* represents the original series needing to be normalized, *max(Series)* represent the maximum data value in the original series, and *Full Value* represents the value that we wish to normalize the original series to.

$$\text{Normalized Series} = \text{Series} \times \left(\frac{\text{Full Value}}{\text{max(Series)}}\right) \quad (2)$$

Then, we merged all the normalized individual writing episodes, estimated magnitude series, entropy series and amplitude series, and sorted them by time stamps. By doing this, we can get combined hand-motion information of all individuals. Subsequently, we calculated the mean, variance, standard deviation and root mean square for these series. For the group series (the group estimated magnitude and entropy), we calculated the mean cross rate, skewness and kurtosis in addition to the features above. We extracted the maximum value, minimum value, mean and mode from the group writing episodes.

Besides the features extracted above, we also wish to know how different the students' hand motions are in the lectures. Thus, we calculated the Euclidean distance between every two-series combination of individuals' estimated magnitude series and entropy series to get a set of distances. We extracted the mean, variance, standard deviation and root mean square from the distance sets.

We also combined the individual heart rate and heart rate variability series into group heart rate and heart rate variability series without normalization, sorting them by time stamps. The normalization process is not required because the individual differences in heart rate activities is not large. We extracted the mean, variation, standard deviation and root mean square from the mixed data. We calculated the mean by using a weighted average, with the confidence values serving as the weights. When the confidence value dropped below 0.6, we discarded the heart rate and heart rate variability data from the other features as the motion artifacts likely have a strong presence. Table III lists all the features.

TABLE III. HIGH-LEVEL FEATURES

Origin	Features Name
Group Writing Episodes	Maximum, Minimum, Mean, Mode
Mixed Individual Writing Episodes	Mean, Variation, Standard Deviation, Root Mean Square
Mixed Individual Amplitude Serires	Mean, Variation, Standard Deviation, Root Mean Square
Mixed Individual Estimated Magnitude Serires	Mean, Variation, Standard Deviation, Root Mean Square
Mixed Individual Entropy Series	Mean, Variation, Standard Deviation, Root Mean Square
Group Estimated Magnitude Series	Mean, Variation, Standard Deviation, Root Mean Square, Mean Cross Rate, skewness and kurtosis
Group Entropy Series	Mean, Variation, Standard Deviation, Root Mean Square, Mean Cross Rate, Skewness and Kurtosis
Distance Set of Individual Estimated Magnitude Serires	Mean, Variation, Standard Deviation, Root Mean Square
Distance Set of Individual Entropy Series	Mean, Variation, Standard Deviation, Root Mean Square
Mixed Individual Heart Rate Series	Mean, Variation, Standard Deviation, Root Mean Square
Mixed Individual Heart Rate Variability Series	Mean, Variation, Standard Deviation, Root Mean Square

D. Feature Selection

Redundant and irrelevant features can negatively impact the performance of classification and its computational efficiency. Therefore, a feature selection technique is required to determine the best subset of features. We used a modified forward selection method. The difference between the method used and regular forward selection [22] is that in every iteration, the method keeps *n* subsets of features with top *n* accuracies rather than only one best subset, because regular forward selection may perform too greedily to determine the best subset. Moreover, in this modified method, we use average accuracy, calculated from the leave-three-out validation, to evaluate subsets of features.

E. Classification

In this work, we investigated four supervised machine learning algorithms: decision tree, nearest neighbor, Naïve Bayes and support vector machine with linear kernel. We used Scikit-learn [23], an open-source machine learning library, for the classification tasks.

III. EXPERIMENTAL RESULTS

A. Experimental Setup and Data Collection

In order to evaluate the proposed approach, we performed an experiment in two normal lectures with a total of 14 topic periods. We recorded the time stamps corresponding to the beginning and end of all 14 topics. Thirty students in the class volunteered to wear the Moto 360 smartwatches on their writing hands during the lectures to collect data at the frequencies of 25 Hz and 12.5 Hz for motion and PPG data, respectively. The investigation was approved by the IRB at Texas A&M University and informed consent was provided to all participants. After each lecture, we surveyed the students and obtained their opinions on their 'interest level' and 'perception of difficulty' towards each topic covered in the lectures. We used this information as the ground-truth labels for the classification algorithms. Fig. 6 shows one example of the survey. Since the surveys were anonymized, every topic may receive a diverse set of opinions in terms of the level of interest or the difficulty level. For example, not all students may find the topic of 'bio-potential signals' interesting or difficult. Therefore, we applied a K-means clustering algorithm to better understand the distribution of the class as a whole. While there are three choices for each cognitive state in the survey, we found that the number of choices for 'not very interested' and 'hard' was very small and therefore we ignored them. As a result, we chose the number of clusters for the K-mean clustering to be two. The 14 data samples that represent topics were labeled as 'interested' or 'not interested' for the interest level, and 'difficult' or 'easy' for the perception of difficulty. After applying the processing to get the set of features described in Section II, we created classifiers leveraging the labels acquired from the clustering technique.

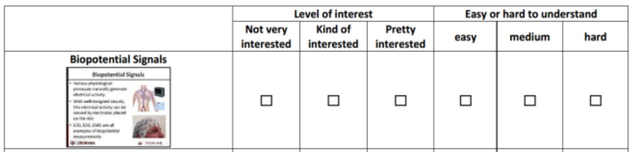

Figure 6. Example of Survey

B. Feature Selection Results

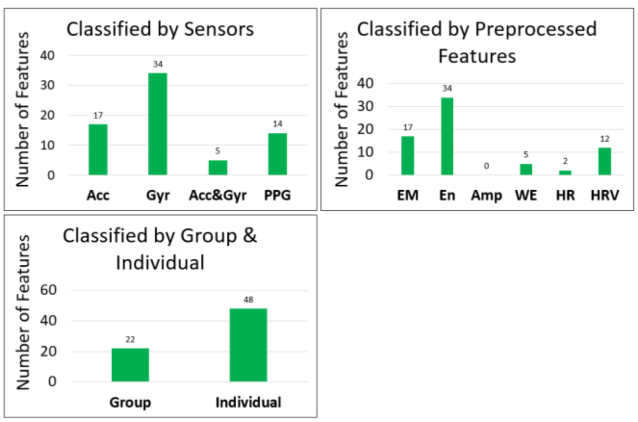

Figure 7. Analysis of Results of Feature Selection

We divided the lectures into 14 episodes, one episode per topic, which means there were 14 labeled data samples in total. The leave-three-out method produced 364 combinations of training and testing sets. We used the mean of the accuracies as the indication to evaluate the subset of features. Table IV shows the results of number of features and the corresponding accuracy of each classifier for both cognitive states. We observed that the number of features that the decision tree algorithm requires is the smallest, and support vector machine and Naïve Bayes require more features. Fig. 7 shows additional statistics and interesting observations: 1) gyroscope data has more importance than accelerometer data and PPG data. This may be because that hand motion has a stronger relation with student cognitive states and the gyroscope data describes the hand motion more effectively. 2) The heart rate variability is more helpful than the pure heart rate. 3) The amplitude has no contribution to the results, but estimated magnitude, calculated from amplitude, offers valuable insights, which may be because the amplitude is too noisy. 4) The features related to entropy play a principal role, which means entropy can help extract useful information from raw data. 5) Features from the individual series are more important than the group, which may be due to the normalization.

TABLE IV. RESULTS OF NUMBER OF FEATURES AND ACCURACY

Classifier	Cognition	Number of Features	Accuracy
Decision Tree	Interest	7	98.99%
	Difficulty	5	93.50%
NN	Interest	8	97.07%
	Difficulty	10	88.29%
SVM	Interest	14	98.08%
	Difficulty	10	95.79%
Naive Bayes	Interest	9	93.22%
	Difficulty	8	91.85%

Correlation-based feature selection [22] is another popular feature selection method, which is model independent. To better understand which features contain more useful information, we calculated the correlation of each feature and the classifying labels and ranked them. Table V shows four features with the highest correlations for both 'interest level' and 'perception of difficulty'. Due to the lack of space, the entire table is not included. The table demonstrates that the hand motion and heart activities are more related to interest level than perception of difficulty. Moreover, most of the features with high correlation are selected in our modified forward selection method, which validates our feature selection approach.

TABLE V. COREELATION RANKING OF FEATURES

Cognition	Name of Features	Correlation
Interest	root mean square of gyroscope individual entropy	0.7087
	mean of gyroscope individual entropy	0.7060
	mean of accelerometer individual entropy	0.6846
	mean of gyroscope group entropy	0.6702
Difficulty	standard deviation of individual writing episodes	0.6015
	variance of individual writing episodes	0.5913
	root mean square of gyroscope distance set of individual estimated magnitude	0.5858
	Standard deviation of gyroscope distance set of individual estimated magnitude	0.5692

B. Classification Results

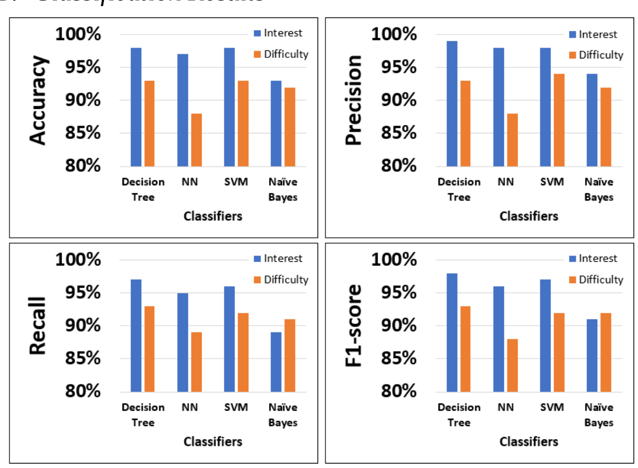

Figure 8. Performance of Classifiers

Fig. 8 shows the performance of each classifier with their own best subset of features. Decision tree and support vector machine provide the best performance, where the best accuracies reach 98.99% for 'interest' prediction and 95.79% for 'difficulty' prediction. For other metrics, including precision, recall and F1-score, all classifiers offer good performance, which demonstrates that the proposed approach is effective. Moreover, the overall results for 'interest level' prediction are better than those for 'perception of difficulty' prediction, which means the students' interest level is more closely related to the students' hand motion and heart activities. This conclusion is consistent with the results of the correlation-based feature ranking.

IV. Conclusion

In this paper, we proposed an approach to model students' cognitive states based on the data from wrist-worn IMUs and PPG sensors. We applied various signal processing techniques to generate a broad set of features. We used a modified forward selection method, combined with leave-three-out validation to get the best subset of features. In order to evaluate the proposed approach, we performed an experiment with 14 topics covered in two lectures with 30 students. Our detection accuracies exhibit the effectiveness of the proposed techniques. For future work, adding additional IMUs can likely lead to additional insights. Leveraging other physiological sensors including EMG, EEG and eyeball tracking could potentially improve the performance of our classifiers.

Acknowledgment

This work was supported in part by TerraSwarm, one of six centers of STARnet, a Semiconductor Research Corporation program sponsored by MARCO and DARPA.

References

[1] R. H. Stevens, T. Galloway, and C. Berka, "Eeg-related changes in cognitive workload, engagement and distraction as students acquire problem solving skills," in *International Conference on User Modeling*. Springer, 2007, pp. 187–196.

[2] P.-S. Tsai, C.-C. Tsai, and G.-H. Hwang, "College studentsâ€™ conceptions of context-aware ubiquitous learning: A phenomenographic analysis," *The Internet and Higher Education*, vol. 14, no. 3, pp. 137–141, 2011.

[3] Y.-M. Huang, Y.-M. Huang, S.-H. Huang, and Y.-T. Lin, "A ubiquitous english vocabulary learning system: Evidence of active/passive attitudes vs. usefulness/ease-of-use," *Computers & Education*, vol. 58, no. 1, pp. 273–282, 2012.

[4] G.-J. Hwang, F.-R. Kuo, P.-Y. Yin, and K.-H. Chuang, "A heuristic algorithm for planning personalized learning paths for context-aware ubiquitous learning," *Computers & Education*, vol. 54, no. 2, pp. 404–415, 2010.

[5] C.-K. Chiou, J. C. Tseng, G.-J. Hwang, and S. Heller, "An adaptive navigation support system for conducting context-aware ubiquitous learning in museums," *Computers & Education*, vol. 55, no. 2, pp. 834–845, 2010.

[6] H.-C. Chu, G.-J. Hwang, and C.-C. Tsai, "A knowledge engineering approach to developing mindtools for context-aware ubiquitous learning," *Computers & Education*, vol. 54, no. 1, pp. 289–297, 2010.

[7] Y.-M. Huang, P.-S. Chiu, T.-C. Liu, and T.-S. Chen, "The design and implementation of a meaningful learning-based evaluation method for ubiquitous learning," *Computers & Education*, vol. 57, no. 4, pp. 2291–2302, 2011.

[8] S.-L. Wang and C.-Y. Wu, "Application of context-aware and personalized recommendation to implement an adaptive ubiquitous learning system," *Expert Systems with applications*, vol. 38, no. 9, pp. 10831–10838, 2011.

[9] C.-M. Chen and Y.-N. Tsai, "Interactive augmented reality system for enhancing library instruction in elementary schools," *Computers & Education*, vol. 59, no. 2, pp. 638–652, 2012.

[10] Á. Di Serio, M. B. Ibáñez, and C. D. Kloos, "Impact of an augmented reality system on students' motivation for a visual art course," *Computers & Education*, vol. 68, pp. 586–596, 2013.

[11] N. A. El Sayed, H. H. Zayed, and M. I. Sharawy, "Arsc: Augmented reality student card," *Computers & Education*, vol. 56, no. 4, pp. 1045–1061, 2011.

[12] H. Salmi, A. Kaasinen, and V. Kallunki, "Towards an open learning environment via augmented reality (ar): Visualising the invisible in science centres and schools for teacher education," *Procedia-Social and Behavioral Sciences*, vol. 45, pp. 284–295, 2012.

[13] R. Chen and X. Wang, "An empirical study on tangible augmented reality learning space for design skill transfer," *Tsinghua Science & Technology*, vol. 13, pp. 13–18, 2008.

[14] C. Conati, "Probabilistic assessment of user's emotions in educational games," *Applied Artificial Intelligence*, vol. 16, no. 7-8, pp. 555–575, 2002.

[15] M. Carroll, C. Kokini, R. Champney, R. Sottilare, and B. Goldberg, "Modeling trainee affective and cognitive state using low cost sensors," in *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, 2011.

[16] B. Woolf, W. Burleson, I. Arroyo, T. Dragon, D. Cooper, and R. Picard, "Affect-aware tutors: recognising and responding to student affect," *International Journal of Learning Technology*, vol. 4, no. 3-4, pp. 129–164, 2009.

[17] T. Dragon, I. Arroyo, B. P. Woolf, W. Burleson, R. El Kaliouby, and H. Eydgahi, "Viewing student affect and learning through classroom observation and physical sensors," in *International Conference on Intelligent Tutoring Systems*. Springer, 2008, pp. 29–39.

[18] M. Raca, L. Kidzinski, and P. Dillenbourg, "Translating head motion into attention-towards processing of student's body-language." *International Educational Data Mining Society*, 2015.

[19] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[20] S. Nemati, M. M. Ghassemi, V. Ambai, N. Isakadze, O. Levantsevych, A. Shah, and G. D. Clifford, "Monitoring and detecting atrial fibrillation using wearable technology," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*. IEEE, 2016, pp. 3394–3397.

[21] P. Stein, PhD and R. Kleiger, MD, "Insights from the study of heart rate variability," *Annual review of medicine*, vol. 50, no. 1, pp. 249–261, 1999.

[22] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825–2830, 2011.